

Hormonal and modality specific effects on males' emotion recognition ability

Adi Lausen^{*1,3}, Christina Broering¹, Lars Penke^{2,3}, Annekathrin Schacht^{1,3}

¹Department of Affective Neuroscience and Psychophysiology, Institute of Psychology, University of Goettingen, 37073 Goettingen, Germany

²Department of Biological Personality Psychology, Institute of Psychology, University of Goettingen, 37073 Goettingen, Germany

³Leibniz ScienceCampus "Primate Cognition", 37077 Goettingen, Germany

*Email: adi.lausen@psych.uni-goettingen.de

Abstract

Successful emotion recognition is a key component of human socio-emotional communication skills. However, little is known about the factors impacting males' accuracy in emotion recognition tasks. This pre-registered study examined potential candidates, focusing on the modality of stimulus presentation, emotion category and individual baseline hormone levels. In an additional exploratory analysis, we examined the association of testosterone x cortisol interaction with recognition accuracy and reaction times. We obtained accuracy and reaction time scores from 282 males who categorized voice, face and voice-face stimuli for nonverbal emotional content. Results showed that recognition accuracy was significantly higher in the audio-visual than in the auditory or visual modality. While Spearman's rank correlations showed no significant association of testosterone (T) with recognition accuracy or with response times for specific emotions, the logistic and linear regression models uncovered some evidence for a positive association between T and recognition accuracy as well as between cortisol (C) and reaction time. In addition, the overall effect size of T by C interaction with recognition accuracy and reaction time was significant, but small. Our results establish that audio-visual congruent stimuli enhance recognition accuracy and provide novel empirical support by showing that the interaction of testosterone and cortisol relates to males' accuracy and response times in emotion recognition tasks.

Keywords: Emotion Recognition, Prosody, Facial Expressions, Testosterone, Cortisol, Dual-hormone hypothesis

Introduction

Emotion recognition is a basic skill thought to carry clear advantages for predicting behaviour, as well as forming and maintaining social bonds (e.g., Fischer and Manstead, 2008; Soto and Levenson, 2009). Conversely, difficulties in correctly recognizing emotional expressions can lead to problematic social relationships and might result in the development and maintenance of psychopathology (e.g., Keltner and Kring, 1998; Marsh and Blair, 2008; Kohler et al., 2009). Intriguingly, research on sex differences highlights that males are less accurate than females when completing emotion recognition tasks (e.g., Hall, 1984; Thompson and Voyer, 2014). However, effect sizes were comparably small and multiple factors known to impact the ability to recognize emotions have yet to be fully controlled for (e.g., Hall et al., 2000; see Chaplin, 2015; Fischer and LaFrance, 2015; Hyde, 2014; Schirmer, 2013 for an overview regarding explanations for sex-based behaviour patterns). Moreover, there is no direct evidence to explain why males have an often-assumed disadvantage in accurately recognizing emotions. Therefore, the main aim of this study was to systematically investigate potential factors that might impact males' ability to recognize emotions.

One of the factors supposed to impact emotion recognition is the modality of stimulus presentation (Hall, 1984). In many everyday situations, judgments about others' emotional states require the integration of information from various sensory modalities making use of different cues such as facial expressions, tone of voice (i.e., prosody), or body language (Klasen et al., 2014). Thus, it has been argued that emotion recognition is a multimodal event (Piwek et al., 2015). Indeed, a growing number of studies pointed out that in emotion recognition tasks the stimuli presented in isolation (i.e., visual or auditory) led to lower accuracy scores and slower response times than the audio-visual presentation of emotional expressions (de Gelder and Vroomen, 2000; Kreifelts et al., 2007; Collignon et al., 2008; Baenziger et al., 2009; Paulmann and Pell, 2011; Jessen et al., 2012). Research on unimodal emotion recognition

1 reported better recognition of emotions from faces than from voices (e.g., Waaramaa, 2017).
2 However, these observations were often contradictory (e.g., Kraus, 2017). Furthermore,
3 previous research in the unimodal domains highlighted that specific emotions were not
4 recognized equally well in the auditory and visual modality. In studies on the vocal channel,
5 participants were faster and most accurate to recognize anger (e.g., Juslin and Laukka, 2003;
6 Cornew et al., 2009; Chronaki et al., 2018), while in studies on facial expressions, happiness
7 was shown to be recognized more accurately and faster than any other emotion (e.g., Elfenbein
8 & Ambady, 2002; Palermo & Coltheart, 2004; Montagne et al., 2007; Williams et al., 2009;
9 Nummenmaa and Calvo, 2015; Wells et al., 2016; Kosonogov and Titova, 2018). Despite these
10 converging patterns, it is not yet possible to make definite claims regarding the advantage of
11 certain emotion categories because, at least within the vocal domain, recognition accuracy (RA)
12 was found to be strongly influenced by the type of stimulus used (see Lausen and
13 Hammerschmidt, 2020, for an overview). Whether the voice is a more reliable source than the
14 face in emotion recognition tasks has been rarely pursued, and results are limited to specific
15 emotions, paradigms, as well as, by a number of methodological differences between studies.
16 Thus, until further evidence regarding RA within specific sensory modalities and emotion
17 categories is provided, the direction of these effects remains an open question.

18 Steroid hormones, including reproductive (i.e., testosterone) and stress-related hormones (i.e.,
19 cortisol) were emphasized as potential predictors of individuals' emotion recognition ability
20 (e.g., Gignell et al., 2019). However, only little is known about the effects of these two
21 hormones on emotion recognition. This gap is surprising since testosterone (T) and cortisol (C)
22 receptors are distributed throughout the nervous system with high concentrations in areas
23 associated with emotion processing, such as the hypothalamus and amygdala (Hakamata et al.,
24 2017; Gignell et al., 2019). T is released as the end-product of the hypothalamic-pituitary-
25 gonadal axis (see Zilioli and Bird, 2017, for details), while C is released by the hypothalamic-

1 pituitary-adrenal (HPA) axis in response to psychological and physical stress or challenges
2 (Jankord and Herman, 2008; see McEwen, 2008, for a comprehensive review on consequences
3 of C on health and diseases). Both T and C have been associated to a range of human social
4 behaviours, including competition, aggression, social status and motivation (e.g., Archer et al.
5 2005; van Bokhoven et al., 2005; Archer, 2006; Cohen et al., 2006; Schultheiss et al., 2016;
6 Carré and Archer, 2017; Zilioli and Bird, 2017). However, only few studies have assessed the
7 association between T or C concentrations on emotion recognition in both sexes, and an even
8 smaller subsection has specifically addressed the impact of these two hormones on males'
9 ability to recognize emotions.

10 The most direct evidence of an association between T and emotion recognition comes from
11 studies showing that in females exogenous administration of T impaired their sensitivity to
12 threatening facial expressions (van Honk and Schutter, 2007), and that a 0.5-mg single-dose T
13 administration altered connectivity to brain areas (e.g., left inferior frontal gyrus) that are
14 associated with the identification and processing of emotional expressions (van Honk et al.,
15 2011; Bos et al., 2016; see also Bos et al., 2012, for a comprehensive review on placebo-
16 controlled single hormone administration studies). In addition, there is evidence that individuals
17 with higher baseline T were less attuned to their partners' thoughts and feelings (e.g., lower
18 empathic accuracy), independent of their sex (e.g., Ronay and Carney, 2013; Nitschke and
19 Bartz, 2020). Briefly summarized, the results across these studies indicate that T has a negative
20 relationship with a variety of empathic processes and suggest that higher levels of T might
21 predict deficits in RA. However, other researchers assessing the causal effect of exogenous T
22 and the moderating effect of prenatal androgenic biomarkers (i.e., 2D:4D digit ratio) on
23 empathy found no robust evidence that T administration or early T exposure causes
24 impairments in males' cognitive empathy (e.g., Nadler et al., 2019).

1 The findings from the few studies on the association between naturally occurring T levels and
2 males' performance accuracy in explicit emotion recognition tasks also yields inconclusive
3 results. By presenting participants with emotional facial expressions at two different intensity
4 levels (i.e., 50% and 100%), Rukavina et al. (2018) found that RA decreased when baseline
5 salivary T was high, especially for full-blown expressions of sadness and for disgust when
6 presented at 50% intensity. Similarly, Fujisawa & Shinohara (2011) reported a negative
7 correlation between baseline salivary T levels and emotion recognition in male adolescents. In
8 contrast, a more recent study, found a positive association between competition-induced
9 changes in T levels and emotion recognition (Vongas and Al Hajj, 2017). An fMRI study by
10 Derntl et al. (2009) found no correlation between blood T levels and males' RA. However, the
11 authors reported an increased amygdala activity in individuals with high T levels during the
12 presentation of fearful and angry faces. Moreover, the authors found a negative correlation
13 between T and RT to fearful male faces (the higher the T level, the faster the response time).
14 Similar patterns were reported by a variety of studies which showed that, through increased
15 amygdala activity, T directs attention to threat related social cues, in particularly to emotions,
16 such as fear (e.g., Bos et al., 2013) and anger (e.g., van Honk et al., 1999; Wirth and Schultheiss,
17 2007). In addition, during approach-avoidance tasks, both healthy and social anxious females
18 showed less avoidance and faster RTs towards angry faces after T administration (e.g., Enter et
19 al., 2016; 2014). Accordingly, it has been argued that increased levels of T reduce the sensitivity
20 towards threat-related emotional expressions. They thereby facilitate approach behaviour as
21 apparent in changes in RTs and amygdala activation, which is strongly linked to the behavioural
22 fight and flight system (Derntl et al. 2009).

23 Concerning the relationship between the processing of emotional material and changes in C
24 levels, the reported evidence across studies is also inconsistent. Although the empirical
25 evidence converges to some extent that in *emotional Stroop*-, *induced stress*- and *cognitive*

tasks, participants with heightened baseline salivary C levels orient their attention towards threatening stimuli, i.e., angry faces (van Honk et al., 2000; Roelofs et al., 2007; Ellenbogen et al., 2010), opposite patterns were also observed (e.g., van Honk et al., 1998). Inconclusive results were also reported on the association between RA and C levels in explicit emotion recognition tasks. For instance, Feeney et al. (2012) found that higher baseline C levels are negatively associated with RA (lower performance) and RTs (faster response times) for angry and happy faces. Conversely, other studies reported that in induced stress contexts RA was positively associated with higher C levels (Deckers et al., 2015) and that this association is moderated by high baseline T, especially when categorizing angry and fearful facial expressions (Bechtoldt and Schneider, 2016). However, in a placebo-controlled, double-blind study Duesenberg et al. (2016) found no evidence for an association between emotion recognition and increases in C levels (after the administration of 10 mg hydrocortisone) in healthy young individuals.

Taken together, the results from the above-mentioned studies provided conflicting evidence on the associations between T or C alone with RA/RTs in explicit emotion recognition tasks. These contradictory findings are likely the result of a number of methodological differences such as insufficient *statistical power* (i.e., most of these studies had small sample sizes), *type of hormone assessments* (e.g., T assessment from blood or saliva; naturally occurring levels vs. exogenous administration; competition- or stress-induced levels), *storage and analyses of hormone samples* (see Schultheiss et al., 2019, for details), the *population in question* (e.g., adolescents, healthy young adults, clinical populations), as well as, the *number of emotions studied* (which range from two to six across studies). Another possible explanation for the discrepancies is that C may constrain T influences on emotion recognition. To reconcile mixed findings on the roles of T and C in human social behaviour, Mehta and Josephs (2010) proposed the dual-hormone hypothesis (DHH). According to this hypothesis, T predicts a wide range of

behaviours, but only under the condition that C concentrations are low. If C concentrations are high, the T-behaviour association is supposed to be attenuated (Carré and Mehta, 2011; Mehta and Prasad, 2015). This hypothesis was supported by several studies, which demonstrated that across different psychological domains the interaction between T and C is associated with empathy, as well as, dominant, status-relevant, risk-taking and antisocial behaviour (see Sarkar et al., 2019, for an overview). However, it should be noted that other studies reported only small effects (e.g., Dekkers et al., 2019; Grebe et al., 2019), null-findings (e.g., Mazur and Booth, 2014), or even reversed patterns [i.e., T related to status-relevant behaviour or facial dominance for high but not low C (e.g., Welker et al., 2014; Kordsmeyer et al., 2018)] for the DHH. Considering the interaction between the HPG and HPA axes might nevertheless lead to more reliable predictions regarding emotion recognition than the assumption of a single-hormone association (Sarkar et al., 2019; Carré and Mehta, 2011).

Against the background of the above-mentioned findings, the present study had three major aims. Firstly, it aimed at examining whether males' RA is influenced by the modality of stimulus presentation. We hypothesized that RA would be better in the audio-visual modality than in the auditory or visual modality (1a), and lower in the visual compared to the auditory modality (1b). Second, we aimed to replicate previous findings by examining the extent RA and RTs vary across discrete emotion categories as a function of modality (e.g., Lambrecht et al., 2014). Specifically, we expected higher accuracy scores and faster RTs for disgusted, fearful and sad expressions in the audio-visual than in both the auditory and the visual modality (2a). We also hypothesized that angry expressions would be identified faster and with higher accuracy in the vocal compared to the facial domain, while we expected the reverse pattern for happy expressions (2b). A third aim was to alleviate some of the methodological flaws of previous research by using a large sample size to examine whether variations in males' ability to recognize emotions are due to baseline T level concentrations. We expected a negative

correlation between T and RA (3a), and that participants with high levels of T would specifically react faster to angry and fearful expressions (3b)¹. In addition, we conducted an exploratory analysis on the associations between baseline C and RA, C and RT, as well as on the relationship between RA or RT and the interaction between T and C levels.

Method

The study was approved by the ethics committee of the Georg-Elias-Mueller-Institute of Psychology (University of Goettingen) and conducted in accordance with the ethical principles formulated in the Declaration of Helsinki (2013). Participants gave informed consent and were reimbursed with course credit or 8 Euros per hour.

Participants

A target sample size of 231 males was determined using an approximate correlation power analysis, Bonferroni-corrected for multiple testing ($r = .25$; $\alpha = .05/20$; $1 - \beta = .80$). To account for possible attrition, the sample size was increased by a minimum of 14%. A total of 312 males (age range 18-36 years; $M_{Age} = 24.3$, $SD = 3.7$) were recruited on the university campus using flyers and via the participant database of the Institute of Psychology (*ORSEE*, www.orsee.org), as well as by posts on the social media site *Facebook* and the online platform *Stellenwerk Jobportal* University Goettingen (www.stellenwerk-goettingen.de). Of the 312 recruited subjects, 30 participants were excluded from analysis due to self-reported hearing problems, psychiatric or neurological disorders, or intake of psychotropic/hormone medication. After these exclusions, a total of 282 participants with a mean age of 24.3 years ($SD = 3.8$) were included in the analysis.

¹ All hypotheses tested in the current paper have been pre-registered (osf.io/w2tgr). This pre-registration contained further hypotheses that are not part of the present paper. As in the pre-registration for the exploratory analysis any information regarding decisions/analysis plan is optional we had not provided any information related to this.

Stimulus material

Stimuli were displayed under three experimental modality conditions: auditory, visual and audio-visual. In each experimental condition, stimuli were presented in one of the emotions of interest (i.e., anger, disgust, fear, happiness, sadness) as well as in a neutral state (i.e., baseline expression).

Audio stimuli

The audio stimuli consisted of pseudo-speech (i.e., pseudo-words, pseudo-sentences) and non-verbal vocalizations (i.e., affect bursts). We decided to use pseudo-speech (i.e., a language devoid of meaning) and non-verbal vocalizations as they have been argued to capture the pure effects of emotional prosody, independent of lexical-semantic cues, and thus to be an ideal tool for investigating the expression of emotional information when there is no concurrent verbal information present (Banse and Scherer, 1996; Pell et al., 2015). The stimuli were sampled from well-established databases or provided by researchers, who developed their own stimulus materials. We validated all stimuli in a previous study (Lausen and Schacht, 2018; cf. Lausen and Hammerschmidt, 2020) and selected only a subset of stimuli (i.e., with the highest accuracy) from each database (see **Table 1**).

Table 1 | Description of audio materials

Database	Speakers	Emotions	Nature of material	Number of stimuli selected	Total stimuli
<i>Magdeburg Prosody Corpus</i> (Wendt & Scheich, 2002)	2 actors (1 male/1 female)	Anger, disgust, fear, happiness, sadness, neutral	Pseudo-words	4	48
<i>Paulmann Prosodic Stimuli</i> (Paulmann & Kotz, 2008; Paulmann et al., 2008)	2 actors (1 male/1 female)		Pseudo-sentences	4	48
<i>Montreal Affective Voices</i> (Belin et al., 2008)	8 actors (4 male/4 female)		Affect bursts		48

The physical volume of stimulus presentations across the nine laptops used in the experiment was controlled by measuring sound volume of the practice trials with a professional sound level meter, *Nor140* (Norsonic, 2010, Lierskogen, Norway). No significant difference in volume intensity was observed [$F_{(8,40)} = 1.546, p = 0.173$].

Visual stimuli

Visual stimuli consisted of 24 frontal face photographs (12 males/12 females), extracted from the *Radboud Faces Database* (Langner et al., 2010). The presentation duration of the faces was matched to the length of the voice stimuli (i.e., from 319 ms to 4821 ms). A grey ellipsoid mask, ensuring a uniform figure/ground contrast surrounded the stimuli, with only the internal area of the face visible (9x14 cm, width and height). The stimuli were presented in colour and corrected for luminance across emotion conditions [$F_{(5,137)} = 0.200, p = 0.962$], using *Adobe Photoshop CS6* (Version 13.0.1, 2012, San Jose, CA).

Audio-visual stimuli

In the audio-visual condition, the voice stimuli were simultaneously presented with the face stimuli. Using *Adobe Premiere Pro CS6* (Version 6.0.5) videos were created, matching face and voice stimuli for sex and emotion category.

Procedure, experimental task and saliva samples

Participants were informed that the study required them to provide two saliva samples over a period of about two hours. A day before the main experiment, they were sent an email instructing them to abstain from sports and the consumption of alcohol, drugs or unnecessary medication on the day of the study. Furthermore, they were instructed not to consume drinks containing caffeine within three hours of the experiment and to refrain from eating, drinking (except water), smoking and brushing their teeth within one hour of the experiment. Adherence to these instructions was assessed using a screening questionnaire (the items were adapted from Schultheiss and Stanton, 2009, and are displayed in *Table S1* in supplementary material). Baseline testosterone levels have been demonstrated to be influenced by a 24-hour circadian cycle in which they are highest in the early morning and drop over the course of the day (Dabbs, 1990). As individual differences in peak hormone levels measured in the morning have been argued to be a better predictor of behavioural responses to emotional stimuli than measurements

1 later in the day (Schultheiss and Stanton, 2009), the designated time slot for testing was between
2 9:00am to 11:00am.

3 Participants were tested in groups of up to nine individuals. On the day of the study, after
4 completing the consent form, participants received oral and written instructions about the
5 procedure of the experiment and the collection of saliva samples. The saliva samples were
6 collected before (T1) and after (T2) the *Emotion Recognition Task*². The experiment was
7 programmed using *Python* (Version 2.7.0, Python Software Foundation, Beaverton, OR) and
8 run on a *Dell Latitude E5530 Laptop* with a 15.6 LCD display screen. The audio stimuli were
9 presented binaurally via headphones (*Bayerdynamic DT 770 PRO*).

Emotion recognition task

10 The emotion recognition task consisted of three blocks, each block displaying one of the three
11 experimental conditions: auditory, visual, and audio-visual. Each experimental condition
12 contained 144 stimuli. A permutation was applied to randomize the order in which the
13 experimental conditions were presented to the participants. Six different permutations were
14 created, and each permutation was allocated randomly in blocks of six participants. The order
15 of the stimuli within each experimental condition was completely randomized. The audio and
16 visual stimuli were matched for duration, sex, and emotion category (see *Table S2* in
17 supplementary material for an example of how the audio and visual stimuli were matched).
18 Before each experimental condition, participants were familiarized with the task in a short
19 training session comprised of three stimuli. Each trial began with a blank screen followed by a
20 fixation cross. Following the presentation of a stimulus, a circular answer display appeared,

² The data reported in this paper was obtained within the confines of a larger study. The experiment began with a short demographic questionnaire followed by the *Screening Questionnaire* (Schultheiss and Stanton, 2009), *Multi-Motive Grid* [MMG, Sokolowski et al., 2000] and *Positive and Negative Affect Schedule* [PANAS, Breyer & Bluemke, 2016]. Next, the first saliva sample (T1) was taken. After a short break, the *Emotion Recognition Task* ensued, followed by PANAS, and the collection of the second saliva sample (T2). The saliva samples were collected approximately 10 minutes before and after the emotion recognition task. The experiment ended with the completion of *Multifaceted Empathy Test* short form [MET, Dziobek et al., 2008] and *Big Five Inventory* [BFI, Danner et al., 2016]. As MMG, PANAS, MET and BFI are not relevant to the present manuscript they are not further reported.

1 containing all six categories of interest (i.e., anger, disgust, fear, happiness, sadness, neutral)
2 and the selection cursor, which appeared in the centre of the display. The sequence of the
3 emotion labels was randomized for each participant and remained the same throughout the task.
4 Participants had to select an emotion category, using the mouse to move the cursor, before the
5 next stimulus was presented. Reaction times were measured, starting with the onset of the
6 answer display and ending with the participant's response. Although there was no time limit
7 for emotion judgments, participants were instructed to respond as accurately and quickly as
8 possible to the presented stimuli. **Figure 1** displays the time course of the emotion recognition
9 task.

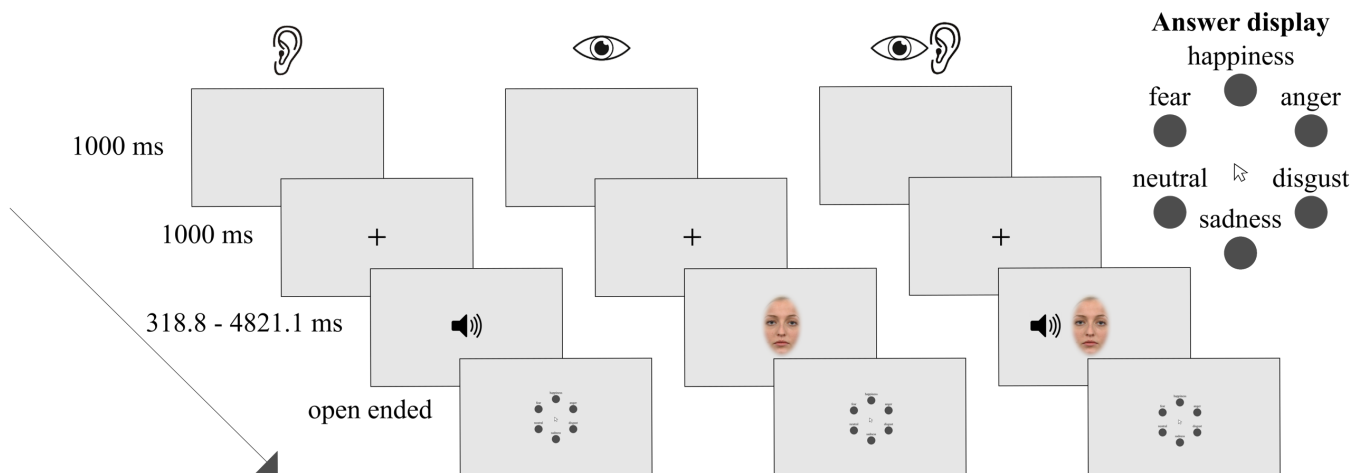


Figure 1 | Emotion recognition task

Each trial began with a blank screen (shown for 1000ms) which was followed by a fixation-cross appearing at the center of the screen (for 1000ms) at which participants were asked to fixate throughout the trial. After the presentation of the stimulus a circular answer display containing all six categories of interest (i.e., anger, disgust, fear, happiness, neutral and sadness) and the selection cursor (which appeared in the center of the display) were presented. The responses were made by using the mouse to move the cursor. Reaction times were measured, starting with the onset of the answer display and ending with the participant's response. There was no time limit for emotion judgments. Participants could hear/see the stimulus only once. The presentation of the stimuli was initiated by pressing the *Spacebar*-key at the beginning of each block. At the end of each block a visual message in the center of the screen instructed participants to take a break if they wished to or to press the *Spacebar*-key to proceed with the next block.

Saliva samples

10 The two saliva samples (2 ml per sample) were collected from each participant via passive drool
11 through a straw (Schultheiss et al., 2012) into an *IBL SaliCap* sampling device. These plastic
12 vials were stored frozen at -80°C until shipment on dry ice to the Endocrinology Laboratory at
13 Technical University of Dresden. At this facility, the samples were analysed for T and C levels

via chemiluminescence immunoassays with high sensitivity (IBL International, Hamburg Germany). The intra- and inter-assay coefficients of variation for T were < 11% and for C < 8%. For T the variance between participants was 14.81% and 3.85% within participants with an intra-class correlation coefficient (ICC) of 79.35%, while for C the variance between participants was 23.78% and 28.20% within participants with an ICC of 45.74%. As the distributions of T and C were positively skewed ($T_{\text{skewness}} = 1.56$; $C_{\text{skewness}} = 1.49$) a log-transformation was performed (e.g., Mehta et al., 2015a; Kordsmeyer et al., 2018). The log-transformation reduced skewness substantially [$\log(T)$ skewness = -0.06; $\log(C)$ skewness = 0.01]. Outliers were winsorized to ± 3 standard deviations (Mehta et al., 2015a). Going beyond the pre-registered analyses, we averaged the two baseline measures for T and C in order to obtain a more reliable value and to cover the observation interval (Idris et al., 2017; Kordsmeyer et al., 2018), as well as to prevent for a “falsely” high or low measurements (Rukavina et al., 2018).

Study design and statistical analysis

A balanced within-subjects factorial design was fitted to assess males’ judgments of emotions. The design was balanced for modalities, emotion categories and encoder sex in each stimulus type. Independent within-participant factors were modalities, emotion categories, stimuli types, and encoder sex. Independent between-participant variables were T and C. Dependent variables were RA and RT.

In line with our preregistration, the primary analysis for our first and second hypotheses was performed using *Friedman-* and *Wilcoxon-rank-sum* tests. For the association between the dependent variables (RA, RT) and T levels we ran *Spearman* correlations (H3a, b). The exploratory analyses of the quantitative variables T and C were performed using generalized linear models (*quasi-binomial logistic regression*) for the binary response variable emotion recognition and linear models for the response variable reaction time, which was normalized

by log transformation. The dispersion parameter of the quasi-binomial model accounted for dependencies caused by repeated measurements within the participants. Modality and emotion category were fitted as nominal variables and stimulus duration³ as quantitative variable. The log-transformed T and C concentrations were mean-centered and multiplied to create an interaction term (Mehta et al., 2015a). Significant interactions were decomposed using the procedures of Aiken & West (1991) and Dawson (2014). Corrections for multiple testing were implemented using Bonferroni's method.

For the descriptive analysis of the data, *relative frequencies*, *confusion matrices* and Wagner's (1993) *unbiased hit rate* (H_u), which is the rate of correctly identified stimuli multiplied by the rate of correct judgments of the stimuli, were calculated. The data was analysed using the R language and environment for statistical computing and graphics version 3.4.3 (R Core Team, 2017) and the integrated environment R-Studio version 1.0.153 (used packages: *coin*; *ggplot2*; *glm*; *MASS*; *pwr*; *rsq*).

Results

Descriptive analysis

Audio-visual emotional expressions were recognized with approximately 90% accuracy (lowest identification rate 89% for disgust). Angry expressions were recognized with better accuracy from the voice (90%) than the face (82%). Conversely, for fearful, happy and sad expressions accuracy scores were higher when presented visually ($85\% \leq \text{accuracy scores} \leq 99\%$) than auditorily ($72\% \leq \text{accuracy scores} \leq 77\%$). Neutral expressions had high accuracy scores in all three conditions of stimulus presentation ($90\% \leq \text{accuracy scores} \leq 95\%$).

³ Research has shown that emotions are not expressed to the same extent in vocal stimuli (e.g., Hammerschmidt and Juergens, 2007; Rigoulot et al., 2013; Schaerlaeken and Grandjean, 2018; Lausen and Hammerschmidt, 2020) and as indicated by our preliminary analysis, the average presentation times were different across the different emotions (see *Table S3* in supplementary material). To control for these differences, we implemented *stimuli duration* as a factor in our models (this decision was also based on our analysis plan from the pre-registration).

1 Participants were faster at recognizing disgust, fear, happy, sad and neutral expressions in the
 2 visual and audio-visual modalities (median (*Md*) values between 1.03 sec. to 1.46 sec.) than in
 3 the auditory modality (*Md* values between 1.50 sec. to 1.95 sec.). Although the RTs for
 4 disgusted, sad and neutral expressions were similar in the visual and audio-visual modalities,
 5 participants were slightly faster at recognizing fear and happy in the visual than audio-visual
 6 modality. For angry expressions, the RTs were much shorter in the audio-visual (1.23 sec.) than
 7 in the auditory and visual modality, but much longer in the visual (1.53 sec.) than in the auditory
 8 modality (1.47 sec.). **Figure 2** illustrates participants' RA (panel A) and RTs (panel B) by
 9 modality and emotion categories.

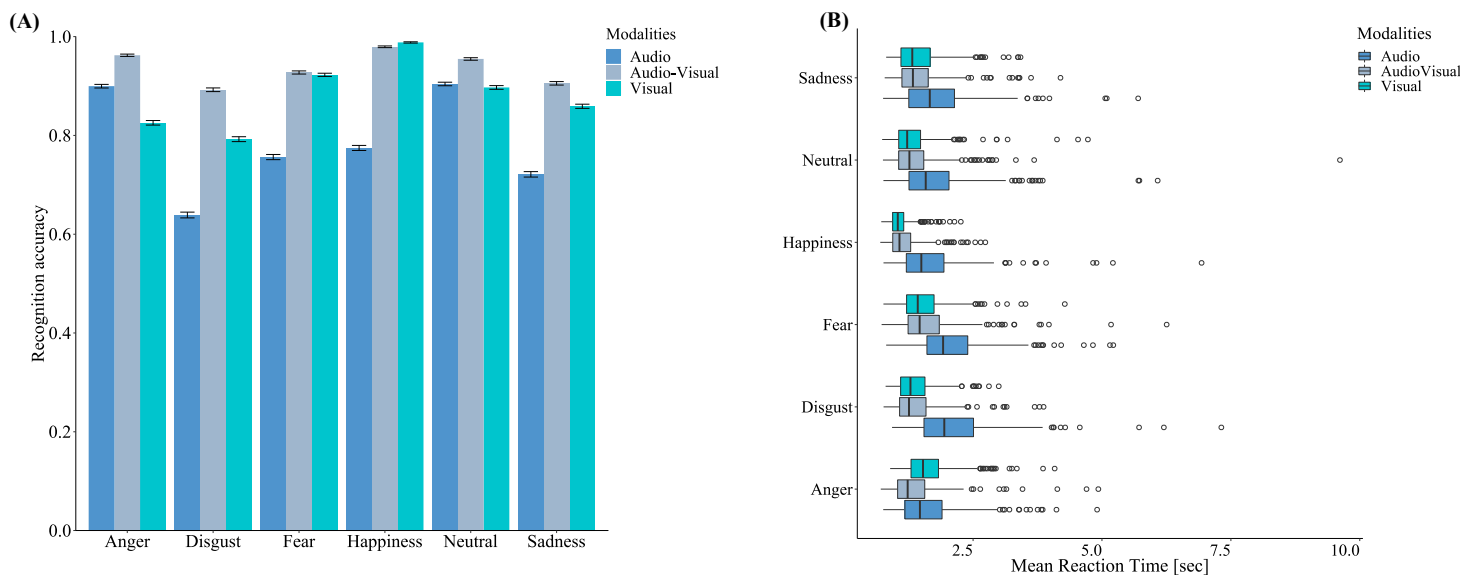


Figure 2 | Recognition accuracy (RA) and reaction times (RTs) by modality and emotion categories
 The bar charts (**panel A**) display RA, while the boxplots (**panel B**) illustrate the mean RT distributions. Error bars represent the standard error. The boxplots indicate that the distributions of RT are right skewed.

10 In all three modalities participants often misclassified happy and sad expressions as neutral. In
 11 the auditory and audio-visual modalities angry was mistaken for fearful, neutral for angry and
 12 fearful for sad. In the visual modality fear was confused with disgust, whereas anger and neutral
 13 were confused with sadness. Participants frequently misclassified disgust with anger in the
 14 visual and audio-visual modalities, while in the auditory modality disgust was mistaken for

1 neutral. The error classification patterns along with the unbiased hit rates are presented in **Table**

2 **2.**

Table 2 | Confusion matrices and unbiased hit rates (H_u) for participants judgments of emotion categories

Modality	Emotions portrayed	Emotion judgments						
		<i>Anger</i>	<i>Disgust</i>	<i>Fear</i>	<i>Happiness</i>	<i>Neutral</i>	<i>Sadness</i>	H_u
Auditory	<i>Anger</i>	89.97	0.87	3.94	2.25	2.59	0.38	.766
	<i>Disgust</i>	5.13	63.89	6.47	4.14	12.04	8.33	.590
	<i>Fear</i>	2.39	2.56	75.62	1.42	6.00	12.01	.621
	<i>Happiness</i>	1.71	0.40	0.22	77.47	19.73	0.47	.665
	<i>Neutral</i>	5.00	0.77	0.92	2.35	90.41	0.55	.549
	<i>Sadness</i>	1.43	0.74	4.95	2.59	18.17	72.12	.554
Visual	<i>Anger</i>	82.55	3.60	2.87	0.09	3.46	7.43	.638
	<i>Disgust</i>	19.03	79.24	0.71	0.19	0.61	0.22	.704
	<i>Fear</i>	0.75	4.17	92.27	0.21	1.08	1.52	.847
	<i>Happiness</i>	0.09	0.03	0.16	98.83	0.88	0.01	.967
	<i>Neutral</i>	2.47	0.22	0.69	1.51	89.72	5.39	.791
	<i>Sadness</i>	1.99	1.98	3.87	0.17	6.09	85.90	.734
Audio-visual	<i>Anger</i>	96.23	0.68	1.34	0.12	1.05	0.58	.860
	<i>Disgust</i>	7.46	89.24	1.02	0.21	1.20	0.87	.858
	<i>Fear</i>	0.58	2.29	92.74	0.13	1.36	2.90	.873
	<i>Happiness</i>	0.07	0.03	0.10	97.95	1.79	0.06	.969
	<i>Neutral</i>	2.51	0.16	0.37	0.52	95.48	0.96	.859
	<i>Sadness</i>	0.81	0.40	2.90	0.13	5.22	90.54	.855
Across all 3 modalities	<i>Anger</i>	89.58	1.72	2.72	0.82	2.36	2.80	.752
	<i>Disgust</i>	10.54	77.46	2.73	1.51	4.62	3.14	.716
	<i>Fear</i>	1.24	3.00	86.88	0.59	2.81	5.48	.780
	<i>Happiness</i>	0.63	0.15	0.16	91.42	7.46	0.18	.864
	<i>Neutral</i>	3.33	0.38	0.66	1.46	91.87	2.30	.710
	<i>Sadness</i>	1.41	1.04	3.91	0.96	9.82	82.86	.709

Note: The accuracy percentage of correctly judged portrayals are given on the main diagonal in boldface type. H_u = the rate of correctly identified stimuli multiplied by the rate of correct judgments of the stimuli. To calculate row, column and overall percentages see *Table S4* in the supplementary material, which display the raw numbers of stimulus trials.

Main analysis

Recognition accuracy in the three modalities [Aim 1]

3 Participants' RA was significantly influenced by the modality of stimulus presentation
4 (Friedman test: $\chi^2_{(2)} = 448.56, p < 0.001$). The results of Wilcoxon-rank-sum test indicated that
5 RA was significantly higher in the audio-visual modality than in the visual ($z = 12.99, p < 0.001$,
6 $95\%CI = [0.052; 0.062]$, effect size (r) = 0.774) or auditory modality ($z = 14.525, p < 0.001$,
7 $95\%CI = [0.146; 0.163]$, $r = 0.865$). Participants' were also significantly more accurate at
8 discriminating emotions when making judgments on visual than on audio stimuli ($z = 13.553$,
9 $p < 0.001$, $95\%CI = [0.090; 0.108]$, $r = 0.807$). **Figure 3** illustrates RA in the three conditions of
10 stimulus presentation.

Emotion specificity and modality [Aim 2]

11 The modality of stimulus presentation across fearful, disgusted and sad expressions
12 significantly influenced participants' RA (Friedman test: $\chi^2_{(2)} = 400.47, p < 0.001$) and RTs

1 (Friedman test: $\chi^2_{(2)} = 208.77, p < 0.001$). Results comparing RA and RTs between modalities
 2 for each emotion category showed that participants were significantly more accurate and faster

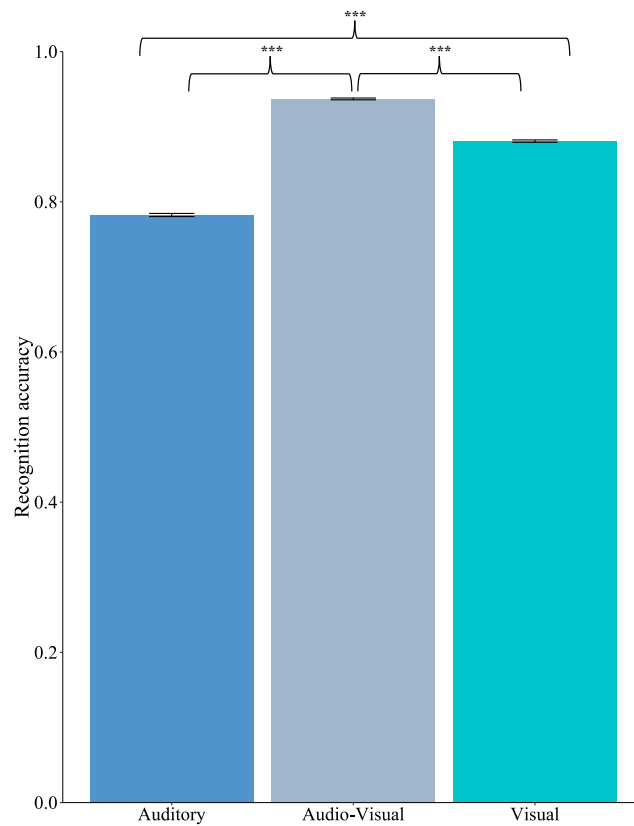


Figure 3 | Bar chart showing recognition accuracy (RA) in the three conditions of stimulus presentation

Error bars represent the standard error. RA was significantly higher for the audio-visual presented stimuli than for the visual or audio stimuli. Accuracy scores were significantly higher for the visual than for auditory modality. All *p-values* were for 3 comparisons (i.e., 3 modalities) Bonferroni corrected.

3 at categorizing these emotions in the audio-visual than auditory modality ($ps < 0.001$; effect
 4 sizes for accuracy ranging from $0.813 < r < 0.852$ and for RTs ranging from $0.422 < r < 0.760$).
 5 Although RA was significantly higher for disgust ($p < 0.001$; $r = 0.605$) and sad expressions (p
 6 < 0.001 ; $r = 0.417$) in the audio-visual than visual modality, the accuracy scores for fear did not
 7 significantly differ between these two modalities ($p = 1.00$; $r = 0.038$). Similarly, we observed
 8 no significant RT differences between the audio-visual and visual modality for these three
 9 emotions ($ps > 0.05$; $0.005 < r < 0.159$). While participants were significantly better at
 10 recognizing angry expressions in the voice than in the face ($p < 0.001$, $r = 0.492$), RTs did not
 11 differ significantly between these two modalities ($p = 1.00$, $r = 0.052$). In contrast, happy,
 12 disgusted, fearful, and sad expressions had significantly higher accuracy scores and faster RTs

- 1 when they were presented visually than auditorily ($ps < 0.001$; $0.625 < r_{\text{Accuracy}} < 0.868$; 0.487
- 2 $< r_{\text{RT}} < 0.816$). **Table 3** displays the test statistics for each modality and emotion category.

Table 3 | Recognition accuracy (RA) and reaction times (RTs) standardized z -scores, p -values, 95% confidence intervals ($CI_{95\%}$) and effect sizes (r) for the comparisons between modalities by emotion categories

Emotions		RA					RT				
		z	p	$CI_{95\%}$		r	z	p	$CI_{95\%}$		r
				LL	UL				LL	UL	
Audio-visual vs. Visual	Anger	13.71	<0.001	0.125	0.146	0.816	-8.645	<0.001	-0.299	-0.200	0.515
	Disgust	10.155	<0.001	0.104	0.125	0.605	0.550	1.00	-0.032	0.569	0.033
	Fear	0.632	1.00	-0.000	0.021	0.038	2.677	0.134	0.019	0.126	0.159
	Happiness	-2.820	0.087	-0.041	-0.000	0.168	3.397	0.012	0.018	0.072	0.202
	Sadness	6.995	<0.001	0.042	0.083	0.417	0.089	1.00	-0.044	0.051	0.005
	Neutral	9.547	<0.001	0.062	0.083	0.568	1.978	0.864	0.000	0.079	0.118
Audio-visual vs. Auditory	Anger	10.579	<0.001	0.063	0.083	0.630	-6.736	<0.001	-0.302	-0.170	0.401
	Disgust	14.315	<0.001	0.250	0.271	0.852	-12.765	<0.001	-0.735	-0.562	0.760
	Fear	13.646	<0.001	0.167	0.188	0.813	-9.653	<0.001	-0.526	-0.366	0.575
	Happiness	14.534	<0.001	0.188	0.208	0.865	-11.709	<0.001	-0.506	-0.373	0.697
	Sadness	13.858	<0.001	0.187	0.208	0.825	-7.087	<0.001	-0.359	-0.208	0.422
	Neutral	8.789	<0.001	0.062	0.083	0.523	-8.659	<0.001	-0.384	-0.242	0.516
Auditory vs. Visual	Anger	8.268	<0.001	0.063	0.104	0.492	-0.865	1.00	-0.094	0.036	0.052
	Disgust	-10.50	<0.001	-0.187	-0.146	0.625	13.711	<0.001	0.597	0.746	0.816
	Fear	-13.318	<0.001	-0.188	-0.167	0.793	12.113	<0.001	0.433	0.579	0.721
	Happiness	-14.574	<0.001	-0.229	-0.188	0.868	13.51	<0.001	0.443	0.571	0.805
	Sadness	-11.603	<0.001	-0.187	-0.146	0.691	8.179	<0.001	0.232	0.370	0.487
	Neutral	0.941	1.00	-0.000	0.021	0.056	10.323	<0.001	0.295	0.420	0.615

Note: The differences in RA and RT between modalities by emotion categories were analyzed using *Wilcoxon-rank-sum test*. All p -values for RA and RT were for 18 comparisons (3 modalities * 6 emotions) Bonferroni corrected. Positive z -scores indicate that RA is higher and RTs longer for the first vs. second modality, whereas negative z -scores indicate that RA is lower and RTs shorter for the first vs. second modality.

Interplay of hormones, recognition accuracy and reaction times [Aim 3]

- 3 Spearman's rank correlation coefficients between T1 and T2 were $r_s = 0.79$ for T and $r_s = 0.60$
- 4 for C. In addition, results showed a negative association between RA and RT ($r_s = -0.26$). No
- 5 significant associations between T and RA ($r_s = 0.03$; $p = 0.64$), T and RT ($r_s = 0.04$; $p = 0.55$),
- 6 C and RA ($r_s = -0.02$; $p = 0.74$) or C and RT ($r_s = 0.11$; $p = 0.06$) were found. *Figure S1* in
- 7 supplementary material illustrates the relationship between T or C and RA/RTs for each and
- 8 across all modalities. Similarly, there were no significant associations between T or C and
- 9 RA/RTs for specific emotion categories (see *Table S5* in supplementary material). The results
- 10 of the logistic and linear regression models indicated small positive main effects of T on RA (β
- 11 $= .028$, $95\%CI = [.039; .047]$, $r_{\text{partial}} = .010$, $p = .004$) and of C on RTs ($\beta = .013$, $95\%CI = [.009$;
- 12 $.016]$, $r_{\text{partial}} = .021$, $p < 0.001$). In addition to the main effects, the results of TxC interaction
- 13 predicting RA and RT are displayed in **Table 4**. Consistent with the DHH, there was a
- 14 statistically significant TxC interaction for RA ($\beta = -.017$, $95\%CI = [-.033, -.001]$, $r_{\text{partial}} = .007$

Table 4 | Logistic and linear regression models of testosterone x cortisol interaction predicting recognition accuracy (RA) and reaction time (RT)

		β	SE	$t_{(121811)}$	p	r_{partial}	95%CI
Quasi-binomial logistic regression (DV = RA)	Testosterone (T)*	0.028	0.010	2.918	0.004	0.010	[0.009; 0.047]
	Cortisol (C)*	0.008	0.010	0.783	0.434	0.003	[-0.011; 0.026]
	T x C*	-0.017	0.008	-2.090	0.037	0.007	[-0.033; -0.001]
Linear regression (DV = RT)	Testosterone (T)*	0.003	0.002	1.781	0.075	0.015	[-0.001; 0.007]
	Cortisol (C)*	0.013	0.002	7.030	<0.001	0.021	[0.009; 0.016]
	T x C*	0.004	0.002	2.531	0.011	0.007	[0.001; 0.007]

Note: β = standardized regression coefficients; *T and C were log-transformed because of skew in the distribution; *Interaction term computed for mean centered predictors. The models were adjusted for stimulus duration, emotions and modalities (see Table S6 in supplementary material); DV = dependent variable.

1 $p = .037$). Simple slopes analyses (**Figure 4**) revealed that T was positively associated with RA
2 when C was low ($b = 0.090$, $se = 0.025$, $t_{(121811)} = 3.685$, $p = 0.002$), while when C was low (b
3 $= 0.090$, $se = 0.025$, $t_{(121811)} = 3.685$, $p = 0.002$), while when C was high, the association between
4 T and RA was not significant ($b = 0.021$, $se = 0.026$, $t_{(121811)} = 0.816$, $p = 0.414$) [**Fig. 4., panel**
5 **A**]. Although not significant, C was positively associated with RA when T was low ($b = 0.050$,

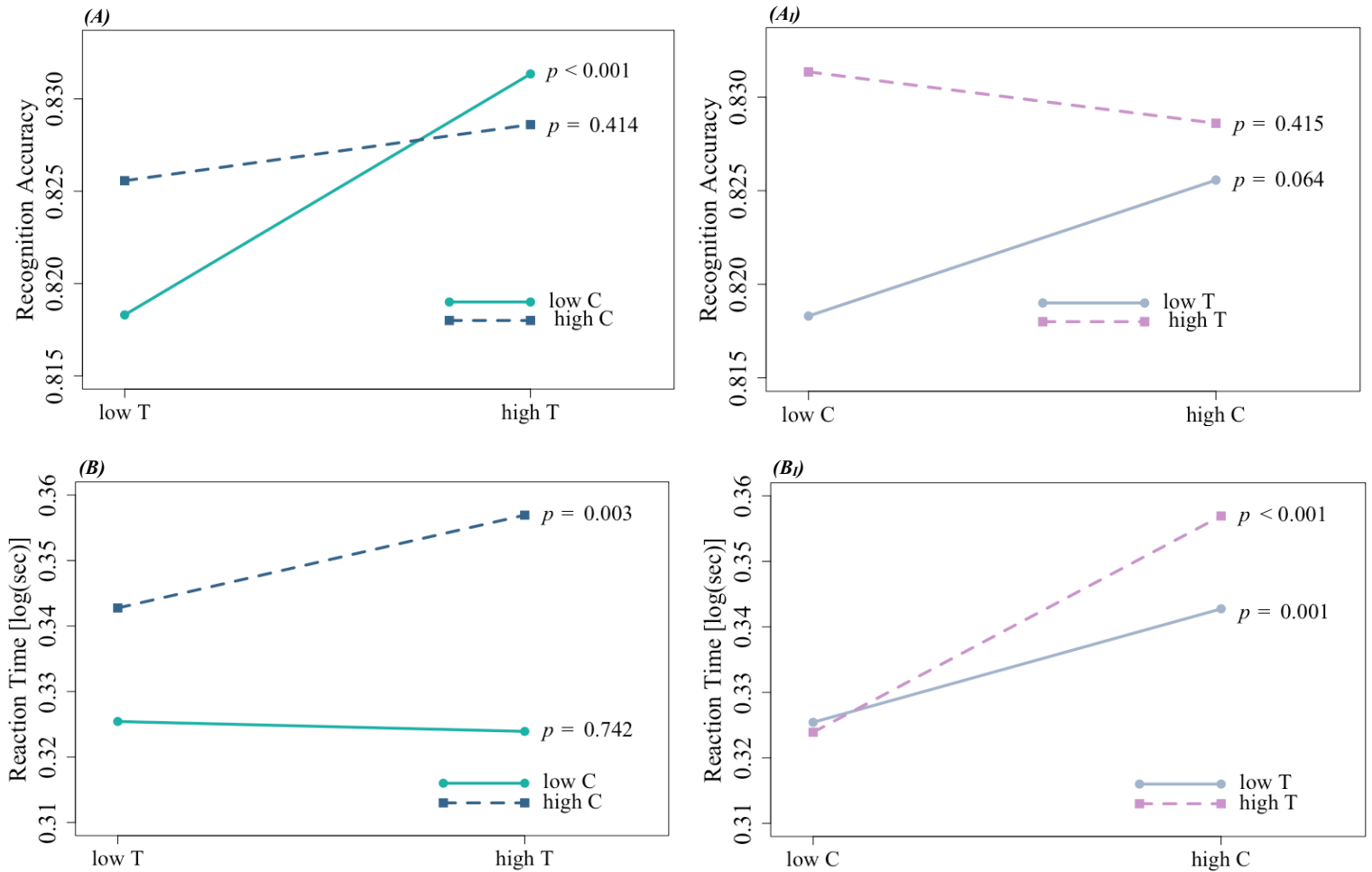


Figure 4 | Recognition accuracy (RA) and Reaction time (RT) as a function of testosterone and cortisol

In all four panels (A, A1, B and B1) the plotted points represent conditional low and high values (± 1 SDs) of testosterone and cortisol.

6 $se = 0.027$, $t_{(121811)} = 1.849$, $p = 0.064$) and negatively associated with RA when T was high (b
7 $= -0.020$, $se = 0.024$, $t_{(121811)} = -0.815$, $p = 0.415$) [**Fig. 4., panel A1**]. Even though results showed

a statistically significant TxC interaction for RT ($\beta = .004$, $95\%CI = [.001, .007]$, $r_{\text{partial}} = .007$, $p = .011$), when decomposing this interaction the simple slope analyses indicated that T was positively associated with RT when C was high ($b = 0.014$, $se = 0.005$, $t_{(121811)} = 2.938$, $p = 0.003$). However, when C was low, the association between T and RA was not significant ($b = -0.002$, $se = 0.005$, $t_{(121811)} = -0.329$, $p = 0.742$) [**Fig. 4., panel B**]. C was positively associated with RT when T was low ($b = 0.017$, $se = 0.005$, $t_{(121811)} = 3.440$, $p = 0.001$) and high ($b = 0.033$, $se = 0.004$, $t_{(121811)} = 7.478$, $p < 0.001$) [**Fig. 4., panel B_I**].

To explore the extent to which TxC interactions on RA and RT were independent of each other, we fitted RT as a covariate in the analysis of RA and RA as a covariate in the analysis of RT. Results showed that RA and RT were strongly associated with each other ($ps < 0.001$) and that TxC interactions on RA and RT were not independent of each other (see *Table S7* in supplementary material).

Discussion

The main objective of the present study was to investigate whether males' RA is influenced by the modality of stimulus presentation in an explicit emotion recognition task. In addition, we examined whether specific emotions are more quickly and accurately detected as a function of modality. Finally, we explored the effects of individual differences in baseline T and baseline C, as well as their interaction with RA and RTs. Our results provide compelling evidence that RA is greatly improved when visual and audio information were jointly presented and that happy expressions were identified faster and with higher accuracy from faces than voices. Conversely, angry expressions were better recognized from voices than faces. Although Spearman's rank correlation coefficients indicated no significant associations between baseline single hormones (i.e., T or C) and RA or RTs, the logistic- and linear regression models uncovered some evidence that there may be a positive association between T and the global

measure of RA as well as a positive association between C and the global measure of RT. Finally, our results showed that TxC interaction was significantly associated with both RA and RTs.

Our data highlights that the audio-visual presentation of emotional expressions significantly contributes to the ease and efficiency with which others' emotions are recognized. This is in line with previous studies showing that the integration of auditorily and visually presented emotional information facilitates emotion recognition (e.g., Baenziger et al., 2009; Paulmann and Pell, 2011; Jessen et al., 2012), reflected in higher accuracy and faster RTs, especially for emotions such as disgust, fear (Collignon et al., 2008) and sadness (Kreifelts et al., 2007). One of the most noticeable differences between the present study and previous investigations was the presentation of several emotions and a neutral category (e.g., Collignon et al., 2008; de Gelder and Vroomen, 2000, included only two emotions) and the measurement of RTs (e.g., not considered in Kreifelts et al., 2007 study). Yet, the facilitation effect concerning stimulus classification manifested for every single emotion category during the audio-visual modality in comparison to the auditory modality. In addition, RA in the audio-visual modality exceeded that of the visual modality for angry, disgusted, neutral and sad emotions, which indicates the comprehensive nature of this integration process. As shown by the present results there are some differences in the effectiveness, with which specific emotions are recognized from voices and faces. Similar to the results reported in a meta-analysis by Elenkin and Ambady (2002), anger was recognized better from voice than faces in our study, while better results for happiness were achieved from the visual compared to the auditory modality. This suggests that sensory modalities do not merely carry redundant information but rather, each may have certain specialized functions for the communication of emotions. Although the estimation of a visual threat (e.g., angry face) can be accurately predicted from close proximity, it has been shown that the louder, higher pitched sound of anger is particularly useful for both, proximal and distal

spaces (see Ceravolo et al., 2016, for details). As it is highly adaptive to recognize and react to a potential threat in the environment (Pichon et al., 2008), the accurate detection of anger might, therefore, rely more on the human auditory than visual system. Previous research on facial expression recognition has consistently reported that happy expressions are recognized more accurately and faster than other basic emotions (e.g., Nummenmaa and Calvo, 2015). Our data provide further support for these findings, but not for our prediction (1b) that emotions communicated by the voice are recognized at higher rates of accuracy than in the visual channel. Nevertheless, it is possible that what determines the recognition advantage of happy faces is not so much their affect, but rather their perceptual and categorical distinctiveness from other emotional expressions (see Calvo et al., 2014, for details) as well as their frequent occurrence in everyday social contexts, thus, tuning the visual system towards efficient recognition of these faces (Nummenmaa and Calvo, 2015). Moreover, it has been argued that physical feature extraction can occur instantaneously for facial expressions, while the interplay of acoustic cues over time occurs in a probabilistic manner (Juslin and Laukka, 2003) and thus, may not engage a similar process for vocal expressions (see Paulmann and Pell, 2011, for details). This could have strengthened the underlying knowledge about emotions leading to improved RA and RTs in the visual modality.

The available evidence regarding the relationship between T and males' emotion recognition ability is by no means clear-cut, making explicit claims about the direction of these effects impossible. The two predictions made in the present study were based on reported observations that in explicit emotion recognition tasks T might have a negative influence on the recognition of emotions (Fujisawa and Shinohara, 2011; Rukavina et al., 2018), and that RTs to threat-related emotional expressions (i.e., angry, fear) would be much shorter with increasing levels of T (Derntl et al., 2009). Similar to other reports in the literature, the results of our primary analysis (i.e., Spearman's rank correlation test) did not provide compelling evidence on the

1 association of single steroid hormones (i.e., T or C) with RA or RTs (Derntl et al., 2009;
2 Duesenberg et al., 2016). In contrast to the reported effect sizes or the significant effects
3 between T and specific emotion categories (Derntl et al., 2009; Rukavina et al., 2018), the
4 correlation coefficients for both hormones were small or close to zero across all modalities in
5 our study. After adjusting for stimuli duration, emotion categories and modalities in our logistic
6 and linear models, results indicated small positive additive effects of T on RA and of C on RTs.
7 The direction of this pattern is similar to those of previous findings which reported that
8 competition-induced T-changes are positively related to the recognition of facial expressions
9 of emotion (Vongas & Al Hajj, 2017). In contrast to other studies which reported higher RA
10 after stress induction (Deckers et al., 2015), our study failed to provide evidence for an
11 association between C-levels and RA. Nevertheless, as indicated by our data high C levels are
12 positively related to RTs (slower response times). This finding, however, is opposite to what
13 Feeney et al. (2012) reported, namely that higher C levels are predictive of faster RTs in
14 emotion recognition tasks. One should note that the association between T and our outcome
15 measures was not in the expected direction (i.e., negative association) and that the C patterns
16 are opposite to what stress induction studies have reported. However, our study design deviated
17 from that of previous studies in theoretical and methodological ways. Studies reporting a
18 negative association between T or C concentrations and emotion recognition either artificially
19 administered T to participants and used exclusively females in their sample (e.g., van Honk and
20 Schutter, 2007; Bos et al., 2016) or, assessed this relationship after stress-induced C changes in
21 different demographic populations [e.g., patients (Deckers et al., 2015); young vs. middle aged
22 adults (Feeney et al., 2012)]. Similar to Vongas and Al Hajj (2017), our participants exhibited
23 normal fluctuating levels of T and thus, it is possible that emotion recognition becomes
24 impaired only at very high levels of T. Due to the ease and simplicity of our emotion recognition
25 task, it is also possible that variance in performance may be greater in competitive interactions

(e.g., Eisenegger et al., 2017; Henry et al., 2017) or in response to laboratory-stress induction methods (e.g., Buchanan and Tranel, 2008; Henckens et al., 2016; Everaerd et al., 2017). Since laboratory methods employed when assessing hormonal changes have been argued to induce a variety of differential hormone responses in individuals (see Welker et al., 2017, for details) one could assume that the small effects for the association between T or C and RA or RTs, respectively, might have been greater/different if we had used similar approaches in our study and examined T and C changes rather than baseline hormone levels.

Most previous research has investigated T and C independently, but neurobiological studies indicated that C may antagonize the relation between T and behaviour (e.g., Chen et al., 1997; Lienen and Josephs, 2010; Goetz et al., 2014). Growing evidence supports that T-related behaviours, such as status seeking, risk-taking and aggression, are better explained by considering the interaction between T and C than by evaluating T fluctuations in isolation (e.g., Popma et al., 2007; Mehta and Josephs, 2010; Carré and Mehta, 2011; Mehta et al., 2015a; but see Grebe et al., 2019). These findings are in line with the idea that environmental stress, as reflected by C concentrations, might reduce or even block the effect of T on direct and indirect behaviours (see Viau, 2002, for details). Following the dual-hormone hypothesis (Mehta and Josephs, 2010), we explored in the present study whether the relationship between baseline T and response variables, namely RA and RT, is enhanced when baseline C levels are low, and attenuated when baseline C levels are high. Similar to the obtained results in the meta-analysis by Dekkers et al. (2019), the overall effect size of T by C interaction on RA and RT was small but significant in our study. When deconstructing this interaction, our results suggested a positive association between T and RA among low C, but not high C, individuals. Further, they indicated positive associations between T and RTs among high C individuals, as well as, between C and RTs among individuals with both high and low T. However, as indicated by the simple slope analysis the association between C and RTs was much stronger for high T- than

1 for low T individuals. In agreement with the dual-hormone model, our results showed that high
2 T was associated with higher RA and faster RTs only when C levels were low. Individuals with
3 high levels of T and C were much slower at categorizing emotions compared to those
4 individuals with high T low C, even though there was little difference in terms of RA. Although
5 our data revealed similarly fast RTs in the emotion recognition task for both low T low C
6 individuals and high T low C individuals, the RA was much lower for low T low C individuals
7 than high T low C individuals.

8 The DHH predicts a negative interaction effect if higher scores on the dependent measures
9 indicate increases in behaviours related to the pursuit of high status (Dekkers et al., 2019;
10 Knight et al., 2020). As a whole, our results are in line with the dual-hormone statistical
11 prediction by showing that individuals with relatively high baseline T and low baseline C
12 performed well on the task in terms of both accuracy and speed (i.e., because in our study the
13 signs for RA (+) and RTs (-) are opposite to what is better performance, DHH predicts a
14 negative interaction effect for RA and a positive interaction effect for RT). This pattern is also
15 consistent with those of prior studies showing that the combination of high T low C is linked
16 to better performance in status-relevant tasks (e.g., Mehta et al., 2015b; Akinola et al., 2016;
17 Henry et al., 2017; Casto et al., 2019). These and other studies (e.g., Mehta and Josephs, 2010;
18 Ponzi et al., 2016; Sherman et al., 2016) suggested that high T low C individuals are prone to
19 enhance their status in different social contexts (e.g., competition, performance, leadership)
20 through their proactive actions as well as assertive, extraverted, confident and dominant
21 behaviours. Henry et al. (2017), for instance, argued that levels of T and C may affect
22 neurotransmission between the amygdala and hypothalamus in response to a threat to status,
23 which then impacts behavioural performance. They suggested that high T and low C may
24 heighten individuals' motivation to perform well in a task (in order to maintain or gain status),
25 while high levels of T and C may disturb individuals' processing of status threat. These

1 explanations can also be extended to our dual-hormone findings. Assuming that high T low C
2 individuals believed that being accurate and fast would be perceived as a marker of high status,
3 it is possible that their motivation to perform well on the task increased. Research has suggested
4 that C-levels tend to increase when individuals appraise a task to be novel and challenging or
5 when experiencing fear of losing social status (e.g., Dickerson, 2008; Sherman and Mehta,
6 2020). Although speculative, the fear of losing social status might have led individuals with
7 high T and high C to compromise RT for RA, thus explaining the difference in RT but not RA
8 in high T individuals with low vs. high C (see, Tops and Boksem, 2011, for effects of C on
9 behavioural inhibition). Individuals with a low T low C profile were argued to engage less in
10 status-related behaviours because they lack a ‘strong power motive’ or ‘dominant personality’
11 (Josephs et al., 2006; Sherman et al., 2016). Therefore, one could speculate that in our study
12 these individuals were less motivated to engage in the task, reflected in their lower recognition
13 accuracy.

14 Although the association between T, C and perceptions of social status is clear in the above-
15 mentioned studies, it remains speculative at this point whether our outcome measures (i.e., RA,
16 RT) or the task-setup are relevant to status. Thus, future studies examining behaviours thought
17 to be relevant to status should preregister their expectations in terms of what outcome would
18 be considered a status-seeking behaviour (Knight et al., 2020).

19 In addition to status-seeking behaviours, evidence suggests that higher levels of T are associated
20 with poorer empathic accuracy (e.g., Nitschke & Bartz, 2020). In a large study ($N = 469$; 323
21 males, 146 females) Zilioli et al. (2015) investigated the association between baseline T,
22 baseline C and empathy. The authors assessed empathy using a well-validated psychometric
23 questionnaire (*Interpersonal Reactivity Index*, IRI) and the *Reading the Mind in the Eyes Test*
24 (RMET). In line with the DHH predictions, they found that baseline T was negatively
25 associated with empathy (as measured by IRI), but only under conditions of low C. However,

1 the dual-hormone effect did not extend to the RMET. In contrast, in our study baseline T-levels
2 were positively associated with emotion recognition. Importantly, this relation was further
3 qualified by low baseline C-levels. There are several possible reasons for why our findings are
4 opposite to what Zilioli et al. (2015) have reported. A possible explanation for the absence of a
5 dual-hormone effect on RMET is related to the low statistical power of their study [while their
6 sample of males was larger than ours by 15%, the measurement accuracy in our study was much
7 higher and the experimental error much lower (282 males x 432 items = 121824 measurements
8 vs. 323 males x 36 RMET items = 11628 measurements)]. Another possible explanation is
9 related to the way the interaction term was modelled [Zilioli et al. (2015) modelled the
10 interaction term by the product of testosterone x logarithm cortisol, while in our study this was
11 modelled by the product of logarithm T x logarithm C]. Other reasons that might explain the
12 discrepancy between our findings and what studies on the association between T, C and
13 empathy have reported are likely related to the psychological constructs measured⁴ or, as
14 mentioned above, to the use of artificially administered T. Thus, it is difficult to tell whether
15 our results on the association between T, C and emotion recognition together with the patterns
16 reported in the literature are due to specific paradigms, experimental procedures (e.g.,
17 controlling for various factors when examining main effects of T or C, see also Grebe et al.,
18 2019) or to the psychological constructs that are measured. Since the patterns we found in our
19 exploratory analysis were not preregistered and the effect sizes for the associations between
20 baseline T, C and TxC interaction with emotion recognition were small, we cannot rule out that

⁴ Although emotion recognition and empathy are often used in research interchangeably due to the fact that they are so interrelated, one important difference between these two constructs is in the way they are measured (for details, see Olderbak and Wilhelm, 2017). While empathy is conceptualized as typical behaviour where individuals are asked about the perception of emotion in general, in others and in oneself (e.g., Batson, 2009), emotion recognition is conceptualized as a maximal effort measure that focus on perceiving emotion in others expressed through certain modalities (e.g., Mayer et al., 2016). Studies that focus on empathy generally involve emotion induction (of the self or another) rather than recognition.

1 it is a false-positive finding. Certainly, more work is needed to test the exact direction and
2 robustness of effects of TxC interaction with RA and RTs in explicit recognition tasks.

3 While our knowledge of how emotional information is integrated and recognized across
4 channels is advancing steadily, the available literature, including the present study, is limited
5 in a number of ways. In comparison to our study, most of the research mentioned above has
6 evaluated a very small number of emotion categories (sometimes as few as two) and did not
7 include a neutral baseline. Further, in some studies the auditory material consisted of speech
8 prosody (words, sentences). This opens up the possibility that the emotional tone of voice
9 interacted with the affective value carried by the sentences'/words' semantic contents. A related
10 issue of past work is the use of emotional exemplars in conflict situations argued to be highly
11 atypical of natural expressions of emotions (Paulmann and Pell, 2011). We addressed these
12 issues by presenting emotion stimuli devoid of meaning (i.e., pseudo-words, pseudo-sentences
13 and affect bursts) which always contained a congruent set of cues (i.e., encoder sex, stimulus
14 length) to express one of five basic emotions or a neutral state. We chose static faces to ensure
15 our experimental conditions of stimulus presentation were compatible with the majority of prior
16 literature. However, this format has been argued to be of less ecological validity compared to
17 dynamic stimuli (Recio et al., 2011; Krumhuber et al., 2013). While this assumption is still
18 subject to some controversy (see Dobs et al., 2018, for details), future studies would benefit
19 from using datasets of more naturalistic stimuli to further increase ecological validity. In our
20 audio-visual condition we used static facial expressions paired with (by its very nature) dynamic
21 vocal expressions. Although in our study, the audio and visual stimuli did not originate from
22 the same event, this cross-modal approach allowed us to investigate whether emotion
23 recognition is facilitated by an enriched stimulus presentation under controlled conditions.

24 Despite our findings not directly informing the multimodal nature of emotion recognition, they
25 are nonetheless consistent with the idea that emotion recognition processes incorporate all

1 available information, possibly in an involuntary manner, leading to systematically higher
2 accuracy rates. However, one needs to bear in mind that our approach represents a strong
3 simplification compared to a dynamic multisensory environment. For instance, research has
4 shown that dynamic facial stimuli are perceived as more intense, realistic and authentic than
5 static expressions (see Krumhuber et al., 2017; 2013, for comprehensive reviews). Given that
6 in natural settings emotional expressions unfold over time (Verduyn et al., 2009) and involve
7 facial action, vocalization and bodily movement (Keltner et al., 2016), the use of dynamic
8 stimulus material in future research would represent a crucial step for our understanding of the
9 multimodal nature of emotion recognition.

10 To date, many psychoneuroendocrinological studies, including the present research, have used
11 immunoassays to measure salivary T and C due to their cost effectiveness and accessibility to
12 many researchers and labs (Miller et al., 2013; Taylor et al., 2015). However, recent papers
13 (e.g., Schultheiss et al., 2019; Prasad et al., 2019) suggest a potential instability in the validity
14 of immunoassays, influenced by matrix interference (e.g., freeze-thaw cycles, storage length
15 and temperature) and antibody performance (e.g., cross-reactivity, type of antibody used).
16 Compared to mass-spectrometry based assessment, immunoassays seem to inflate the estimates
17 of salivary C (Miller et al., 2013) and T, especially at lower concentrations (see Welker et al.
18 2016, for details). Since prior research has documented inconsistent associations of T or C with
19 behaviour, with some studies revealing strong effects and other studies (including ours)
20 revealing weak or null effects, future studies should consider mass-spectrometry based
21 measurement of hormone levels in saliva, which seems to allow for better replicability (Prasad
22 et al., 2019; Roy et al., 2019). Blood and saliva measurements reflect steroid excretion rates for
23 short periods of time and are influenced by a number of factors such as time of the day, food
24 intake prior to sampling or sleep-related factors (Dettenborn et al., 2012). Alternative sampling
25 methods, such as from hair or fingernails were found to be fairly robust to these factors and to

1 provide a cumulative measure representing steroid excretion levels over a longer period of time
2 (Dettenborn et al., 2016; Matas and Koren, 2016). Thus, an interesting avenue for future
3 research would be to assess whether hair or nail samples provide a stronger test of the
4 relationship between baseline levels of T or C and RA in explicit recognition tasks.

5 It must also be noted that there are many variables, in addition to those that were accounted for
6 in this study, which may impact on baseline T or C levels. These include smoking, hours slept
7 and encoder sex. While we asked participants to refrain from smoking before testing, we did
8 not gather data on chronic smoking habits, nor did we gather data on waking hours. Although
9 T levels seem not to fluctuate with regard to the emotional content of faces, it has been reported
10 that both males and females have elevated levels of T after exposure to faces of the opposite
11 sex (Zilioli et al., 2014). These are worthwhile issues which need to be considered in future
12 research. The homogeneous characteristics of our sample (i.e., university students, narrow age
13 range) may show patterns which do not hold for different sociodemographic subgroups. Given
14 the increased focus on study replicability, future studies would benefit from combining datasets
15 of different laboratories with similar outcome measures in order to reduce costs and increase
16 the external validity, reliability and generalizability of findings. The present study provided
17 evidence for differences in both RA and RTs in the three conditions of stimulus presentation
18 and potentially set the stage regarding the association of baseline TxC with these two response
19 variables. It would thus be worthwhile to expand on these findings and examine whether the
20 same holds true for the other sex. This could be done, for instance, by investigating the
21 interaction between oestradiol and cortisol with RA, as previous research showed that high
22 oestradiol is associated with more externalizing behaviours (linked to emotion-recognition
23 difficulties, see Chronaki et al., 2015), but only when cortisol was low (Tackett et al., 2015).

Conclusion

Humans interact socially, often with the help of emotions. Hence, it stands to reason that their accurate recognition plays a pivotal role in regulating social interaction. Since there is the proviso that females are more accurate than males at recognizing emotions, a more precise knowledge of the factors assumed to impact on their ability would help to convey more accurate results. Given that much of our social interactions depend on the successful recognition of emotional information, it is critical to understand how we make use of different sources of emotional information and to identify whether we base emotional inferences on a particular hierarchy of information channels. The results of our study exemplify that as emotional channel availability increases, there is a corresponding increase in how accurately emotional displays are explicitly recognized. Moreover, they provide compelling evidence that particular channels are more effective for recognizing specific emotions and help explain inconsistencies in the past literature by highlighting that in explicit emotion recognition tasks voice-only expressions do not have a RA advantage over face-only or voice-and-face expressions. In addition, they replicate previous findings by establishing that for particular emotion categories RA and RTs vary as a function of modality.

As most of the previous research has focused on the associations between single hormones and facial emotion recognition, the present study uniquely contributes to the literature by providing a systematic examination of the association of baseline T, C and their interaction with RA and RT across different sensory modalities (i.e., auditory, visual and audio-visual). Although for C, as well as for the interaction between T and C, the analyses were exploratory, they might prove of importance for researchers conducting work in this area to gain a more comprehensive understanding of when these effects emerge and when they do not. They may also yield a substantial theoretical payoff by enabling richer and more accurate predictions concerning the kind of outcomes tied to certain hormone level combinations. Crucially, our study advances

knowledge on the neuroendocrinology of emotion recognition in suggesting that T and C work in concert to regulate emotion recognition. As shown by the patterns in our study, adding C to the list of physiological modulators of T release, represents an important step towards a better understanding of how androgens shape social behaviour and ultimately emotion recognition. It hereby paves the way for impactful future research, especially for the effects regarding TxC interaction with RA and RT.

References

- Aiken, L. S., West, S. G., 1991. Multiple Regression: testing and interpreting interactions. Thousand Oaks, CA, US: Sage Publications.
- Akinola, M., Page-Gould, E., Mehta, P. H., Lu, J. G. 2016. Collective hormonal profiles predict group performance. *Proc. Natl. Acad. Sci.* 113, 9774–9779. <https://doi.org/10.1073/pnas.1603443113>
- Archer, J., 2006. Testosterone and human aggression: an evaluation of the challenge hypothesis. *Neurosci. Biobehav. Rev.* 30, 319–345. <https://doi.org/10.1016/j.neubiorev.2004.12.007>
- Archer, J., Graham-Kevan, N., Davies, M., 2005. Testosterone and aggression: A reanalysis of Book, Starzyk, and Quinsey's (2001) study. *Aggression and Violent Behavior* 10, 241–261. <https://doi.org/10.1016/j.avb.2004.01.001>
- Banise, R., Scherer, K. R., 1996. Acoustic profiles in vocal emotion expression. *J. Pers. Soc. Psychol.* 70, 614–636. <http://dx.doi.org/10.1037/0022-3514.70.3.614>
- Baenziger, T., Grandjean, D., Scherer, K., 2009. Emotion recognition from expressions in face, voice, and body: the multimodal emotion recognition test (MERT). *Emotion* 9, 691–704. <https://doi.org/10.1037/a0017088>
- Batson, D. C., 2009. These things called empathy: eight related but distinct phenomena, in Decety, J., Ickes, W. (Eds.), *The social neuroscience of empathy*, Cambridge, MA: MIT Press, pp. 3–16. <https://doi.org/10.7551/mitpress/9780262012973.003.0002>
- Bechtoldt, M. N., Schneider, V. K. 2016. Predicting stress from the ability to eavesdrop on feelings: Emotional intelligence and testosterone jointly predict cortisol reactivity. *Emotion* 16, 815–825. <http://dx.doi.org/10.1037/emo0000134>
- Belin, P., Fillion-Bilodeau, S., Gosselin, F., 2008. The Montreal affective voices: a validated set of nonverbal affect bursts for research on auditory affective processing. *Behav. Res. Methods* 40, 531–539. <http://dx.doi.org/10.3758/BRM.40.2.531>

- 1 Bos, P. A., Hofman, D., Hermans, E. J., Montoya, E.R., Baron-Cohen, S., van Honk, J. 2016.
2 Testosterone reduces functional connectivity during the ‘Reading the Mind in the Eyes’ Test.
3 *Psychoneuroendocrinology* 68, 194–201. <http://dx.doi.org/10.1016/j.psyneuen.2016.03.006>
- 4 Bos, P. A., van Honk, J., Ramsey, N. F., Stein, D. J., Hermans, E. J. 2013. Testosterone
5 administration in women increases amygdala responses to fearful and happy faces.
6 *Psychoneuroendocrinology* 38, 808–817. <http://dx.doi.org/10.1016/j.psyneuen.2012.09.005>
- 7 Bos, P.A., Panksepp, J., Bluthé, R.M., van Honk, J. 2012. Acute effects of steroid hormones
8 and neuropeptides on human social-emotional behavior: a review of single administration
9 studies. *Front. Neuroendocrinol.* 33, 17–35. <http://dx.doi.org/10.1016/j.yfrne.2011.01.002>
- 10 Breyer, B. Bluemke, M., 2016. Deutsche Version der Positive and Negative Affect Schedule
11 PANAS (GESIS panel), in: *Zusammenstellung Sozialwissenschaftlicher Items und Skalen*.
12 <https://doi.org/10.6102/zis242>
- 13 Buchanan, T. W., Tranel, D., 2008. Stress and emotional memory retrieval: Effects of sex and
14 cortisol response. *Neurobiol. Learn. Mem.* 89, 134–141.
15 <http://dx.doi.org/10.1016/j.nlm.2007.07.003>
- 16 Calvo, M. G., García, A. G., Martín, A. F., Nummenmaa, L., 2014. Recognition of facial
17 expressions of emotion is related to their frequency in everyday life. *J. Nonverbal Behav.* 38,
18 549–597. <https://doi.org/10.1007/s10919-014-0191-3>
- 19 Carré, J. M., Archer, J. 2017. Testosterone and human behavior: the role of individual and
20 contextual variables. *Curr. Opin. Psychol.* 19, 149–153.
21 <http://dx.doi.org/10.1016/j.copsyc.2017.03.021>
- 22 Carré, J. M., Mehta, P. H., 2011. Importance of considering testosterone–cortisol interactions
23 in predicting human aggression and dominance. *Aggress. Behav.* 37, 489–491.
24 <https://doi.org/10.1002/ab.20407>
- 25 Casto, K. V., Hamilton, D. K., Edwards, D. A., 2019. Testosterone and cortisol interact to
26 predict within-team social status hierarchy among Olympic-level women athletes. *Adapt.*
27 *Human Behav. Physiol.* 5, 237–250. <https://doi.org/10.1007/s40750-019-00115-2>
- 28 Ceravolo, L., Fröhholz, S., Grandjean, D., 2016. Proximal vocal threat recruits the right voice-
29 sensitive auditory cortex. *Soc. Cogn. Affect. Neurosci.* 11, 793–802.
30 <https://doi.org/10.1093/scan/nsw004>
- 31 Chaplin, T. M., 2015. Gender and emotion expression: a developmental contextual perspective.
32 *Emot. Rev.* 7, 14–21. <https://doi.org/10.1177/1754073914544408>
- 33 Chen, Sy., Wang, J., Yu, Gq., Liu, W., Pearce, D., 1997. Androgen and glucocorticoid receptor
34 heterodimer formation. A possible mechanism for mutual inhibition of transcriptional activity.
35 *J. Biol. Chem.* 30, 14087–14092. <https://doi.org/10.1074/jbc.272.22.14087>
- 36 Chronaki, G., Wigelsworth, M., Pell, M.D., Kotz, S.A., 2018. The development of cross-
37 cultural recognition of vocal emotions during childhood and adolescence. *Sci. Rep.* 8, 8659.
38 <https://doi.org/10.1038/s41598-018-26889-1>

- 1 Chronaki, G., Garner, M., Hadwin, J.A., Thompson, M. J. J., Chin, C. Y., Sonuga-Barke, E. J.
2 S., 2015. Emotion-recognition abilities and behavior problem dimensions in preschoolers:
3 evidence for a specific role for childhood hyperactivity. *Child Neuropsychol.* 21, 25–40.
4 <https://doi.org/10.1080/09297049.2013.863273>
- 5 Cohen, S., Doyle, W.J., Baum, A., 2006. Socioeconomic status is associated with stress
6 hormones. *Psychosom. Med.* 68, 414–420.
7 <https://doi.org/10.1097/01.psy.0000221236.37158.b9>
- 8 Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., Lepore, F.,
9 2008. Audio-visual integration of emotion expression. *Brain Res.* 1242, 126–135.
10 <https://doi.org/10.1016/j.brainres.2008.04.023>
- 11 Cornew, L., Carver, L., Love, T., 2009. There's more to emotion than meets the eye: a
12 processing bias for neutral content in the domain of emotional prosody. *Cogn. Emot.* 24, 1133–
13 1152. <https://doi.org/10.1080/02699930903247492>
- 14 Dabbs, J. M., 1990. Salivary Testosterone Measurements: Reliability across Hours, Days,
15 Weeks. *Physiol. Behav.* 48, 83–86. [http://dx.doi.org/10.1016/0031-9384\(90\)90265-6](http://dx.doi.org/10.1016/0031-9384(90)90265-6)
- 16 Danner, D. et al., 2016. Die deutsche version des Big Five Inventory 2 (BFI-2), in
17 *Zusammenstellung sozialwissenschaftlicher items und skalen.* <https://doi.org/10.6102/zis247>
- 18 Dawson, J. F., 2014. Moderation in management research: what, why, when, and how. *J. Bus.*
19 *Psychol.* 29, 1– 19. <https://doi.org/10.1007/s10869-013-9308-7>
- 20 Deckers, J.W.M., Lobbestael, J., van Wingen, G. A., Roy P.C. Kessels, A. A., Egger, J. I. M.,
21 2015. The influence of stress on social cognition in patients with borderline personality
22 disorder. *Psychoneuroendocrinology* 52, 119–129.
23 <http://dx.doi.org/10.1016/j.psyneuen.2014.11.003>
- 24 de Gelder, B., Vroomen, J. 2000. The perception of emotions by ear and by eye. *Cogn. Emot.*
25 14, 289–311. <https://doi.org/10.1080/026999300378824>
- 26 Dekkers, T. J., Agelink van Rentergem, J. A., Meijera, B., Popmab, A., Wagemakera, E.,
27 Huizengaa, H. M., 2019. A meta-analytical evaluation of the dual-hormone hypothesis: does
28 cortisol moderate the relationship between testosterone and status, dominance, risk taking,
29 aggression, and psychopathy? *Neurosci. Biobehav. Rev.* 96, 250–271.
30 <https://doi.org/10.1016/j.neubiorev.2018.12.004>
- 31 Derntl, B., Windischberger, C., Robinson, S., Kryspin-Exner, I., Gur, R. C., Moser, E., Habel,
32 U., 2009. Amygdala activity to fear and anger in healthy young males is associated with
33 testosterone. *Psychoneuroendocrinology* 34, 687–693.
34 <https://doi.org/10.1016/j.psyneuen.2008.11.007>
- 35 Dettenborn, L., Tietze, A., Kirschbaum, C., Stalder, T., 2016. Increased hair testosterone but
36 unaltered hair cortisol in female patients with borderline personality disorder.
37 *Psychoneuroendocrinology* 71, 176–179. <http://dx.doi.org/10.1016/j.psyneuen.2016.05.026>

- 1 Dettenborn, L., Tietze, A., Kirschbaum, C., Stalder, T., 2012. The assessment of cortisol in
2 human hair: associations with sociodemographic variables and potential confounders. *Stress*
3 *15*, 578–588. <https://doi.org/10.3109/10253890.2012.654479>
- 4 Dickerson, S. S., 2008. Emotional and physiological responses to social-evaluative threat. *Soc.*
5 *Personal. Psychol. Compass* 2, 1–17. <https://doi.org/10.1111/j.1751-9004.2008.00095.x>
- 6 Dobs, K., Bühlhoff, I., Schultz, J., 2018. Use and usefulness of dynamic face stimuli for face
7 perception studies – a review of behavioral findings and methodology. *Front. Psychol.* 9, 1355.
8 <https://doi.org/10.3389/fpsyg.2018.01355>
- 9 Duesenberg, M., Weber, J., Schulze, L., Schaeuffele, C., Roepke, S., Regen, J. H., Otte, C.,
10 Wingenfeld, K., 2016. Does cortisol modulate emotion recognition and empathy?
11 *Psychoneuroendocrinology* 66, 221–227. <https://doi.org/10.1016/j.psyneuen.2016.01.011>
- 12 Dziobek, I., Rogers, K., Fleck, S., Bahnemann, M., Heekeren, H. R., Wolf, O. T., Convit, A.,
13 2008. Dissociation of cognitive and emotional empathy in adults with Asperger syndrome using
14 the multifaceted empathy test (MET). *J. Autism Dev. Disord.* 38, 464–473.
15 <https://doi.org/10.1007/s10803-007-0486-x>
- 16 Eisenegger C., Kumsta, R., Naef, M., Gromoll, J., Heinrichs, M., 2017. Testosterone and
17 androgen receptor gene polymorphism are associated with confidence and competitiveness in
18 men. *Horm. Behav.* 92, 93–102. <http://dx.doi.org/10.1016/j.yhbeh.2016.09.011>
- 19 Elfenbein, H. A., Ambady, N. 2002. On the universality and cultural specificity of emotion
20 recognition: a meta-analysis. *Psychol. Bull.* 128, 203–235. [https://doi.org/10.1037/0033-](https://doi.org/10.1037/0033-2909.128.2.203)
21 [2909.128.2.203](https://doi.org/10.1037/0033-2909.128.2.203)
- 22 Ellenbogen, M. A., Carson, R. J., Pishva, R., 2010. Automatic emotional information
23 processing and the cortisol response to acute psychosocial stress. *Cognit. Affect. Behav.*
24 *Neurosci.* 10, 71–82. <https://doi.org/10.3758/CABN.10.1.71>
- 25 Enter, D., Spinhoven, P., Roelofs, K., 2016. Dare to approach: single dose testosterone
26 administration promotes threat approach in patients with social anxiety disorder. *Clin. Psychol.*
27 *Sci.* 4, 1073–1079. <https://doi.org/10.1177/2167702616631499>
- 28 Enter, D., Spinhoven, P., Roelofs, K., 2014. Alleviating social avoidance: Effects of single dose
29 testosterone administration on approach–avoidance action. *Horm. Behav.* 65, 351–354.
30 <https://doi.org/10.1016/j.yhbeh.2014.02.001>
- 31 Everaerd, D., Klumpers, F., Voshaar, R. O., Fernández, G., Tendolkar, I., 2017. Acute stress
32 enhances emotional face processing in the aging brain. *Biol. Psychiatry Cogn. Neurosci.*
33 *Neuroimaging* 2, 591–598. <https://doi.org/10.1016/j.bpsc.2017.05.001>
- 34 Feeney, J., Gaffney, P., O'Mara, S. M., 2012. Age and cortisol levels modulate judgment of
35 positive and negative facial expressions. *Psychoneuroendocrinology*, 37, 827–835.
36 <https://doi.org/10.1016/j.psyneuen.2011.09.015>

- 1 Fischer, A. H., LaFrance, M. 2015. What drives the smile and the tear: why women are more
2 emotionally expressive than men. *Emot. Rev.* 7, 22–29.
3 <https://doi.org/10.1177/1754073914544406>
- 4 Fischer, A. H., Manstead, A. S. R., 2008. Social functions of emotion, in Lewis, M., Haviland-
5 Jones, J., Barrett L. F. (Eds.), *Handbook of emotions* (3rd Ed). Guilford Press, New York, pp.
6 456–468.
- 7 Fujisawa, T. X., Shinohara, K., 2011. Sex differences in the recognition of emotional prosody
8 in late childhood and adolescence. *J. Physiol. Sci.* 61, 429–435. [https://doi.org/10.1007/s12576-](https://doi.org/10.1007/s12576-011-0156-9)
9 [011-0156-9](https://doi.org/10.1007/s12576-011-0156-9)
- 10 Gignell, M., Hornung, J., Derntl, B., 2019. Emotional processing and sex hormones, in
11 Schultheiss O. C., Mehta P. H. (Eds.), *Routledge International Handbook of Social*
12 *Neuroendocrinology* (1st Ed.). Routledge, Abingdon, UK, pp. 403–419.
13 <https://www.routledgehandbooks.com/doi/10.4324/9781315200439-24>
- 14 Goetz, S.M.M., Tang, L., Thomason, M. E., Diamond, M.P., Hariri, A.R., Carré, J.M., 2014.
15 Testosterone rapidly increases neural reactivity to threat in healthy men: a novel two-step
16 pharmacological challenge paradigm. *Biol. Psychiatry* 76, 324–331.
17 <http://dx.doi.org/10.1016/j.biopsych.2014.01.016>
- 18 Grebe, N.M., Del Giudice, M., Thompson, M.E., Nickels, N., Ponzi, D., Zilioli, S., Mastrapieri,
19 D., Gangestad, S.W., 2019. Testosterone, Cortisol, and Status-Striving Personality Features: A
20 Review and Empirical Evaluation of the Dual Hormone Hypothesis. *Horm. Behav.* 109, 25–37.
21 <https://doi.org/10.1016/j.yhbeh.2019.01.006>
- 22 Hakamata, Y., Komi, S., Moriguchi, Y., Izawa, S., Motomura, Y., Sato, E., et al., 2017.
23 Amygdala-centred functional connectivity affects daily cortisol concentrations: a putative link
24 with anxiety. *Sci. Rep.* 7, 8313. <https://doi.org/10.1038/s41598-017-08918-7>
- 25 Hammerschmidt, K., Juergens, U., 2007. Acoustical correlates of affective prosody. *J. Voice*
26 21, 531–540. <https://doi.org/10.1016/j.jvoice.2006.03.002>
- 27 Hall, J. A., Carter, J. D., Horgan, T. G., 2000. Gender differences in the nonverbal
28 communication of emotion, in Fischer A. H. (Ed.), *Gender and Emotion: Social Psychological*
29 *Perspectives*, Cambridge University Press, Cambridge, pp. 97–117.
- 30 Hall, J. A., 1984. *Non-Verbal Sex Differences. Communication Accuracy and Expressive Style.*
31 John Hopkins Press, London.
- 32 Henckens, M. J., Klumpers, F., Everaerd, D., Kooijman, S. C., van Wingen, G. A., Fernández
33 G., 2016. Interindividual differences in stress sensitivity: basal and stress-induced cortisol
34 levels differentially predict neural vigilance processing under stress. *Soc. Cogn. Affect.*
35 *Neurosci.* 11, 663–673. <http://dx.doi.org/10.1093/scan/nsv149>
- 36 Henry A., Sattizahn, J. R., Norman, G. J., Beilock, S. L., Maestripieri, D., 2017. Performance
37 during competition and competition outcome in relation to testosterone and cortisol among
38 women. *Horm. Behav.* 92, 82–92. <http://dx.doi.org/10.1016/j.yhbeh.2017.03.010>

- 1 Hyde, J. S., 2014. Gender similarities and differences. *Annu. Rev. Psychol.* 65, 373–398.
2 <https://doi.org/10.1146/annurev-psych-010213-115057>
- 3 Idris, F. P., Wan, Y., Zhang, X., Punyadeera, C., 2017. Within-Day Baseline Variation in
4 Salivary Biomarkers in Healthy Men. *OMICS: A Journal of Integrative Biology* 21, 74–80.
5 <http://dx.doi.org/10.1089/omi.2016.0168>
- 6 Jankord, R., Herman, J. P., 2008. Limbic regulation of hypothalamo-pituitary-adrenocortical
7 function during acute and chronic stress. *Ann. N Y Acad. Sci.* 1148, 64–73.
8 <http://dx.doi.org/10.1196/annals.1410.012>
- 9 Jessen, S., Obleser, J., Kotz, S. A., 2012. How bodies and voices interact in early emotion
10 perception. *PLoS ONE* 7, e36070. <http://dx.doi.org/10.1371/journal.pone.0036070>
- 11 Josephs, R. A., Sellers, J. G., Newman, M. L., Mehta, P. H., 2006. The mismatch effect: when
12 testosterone and status are at odds. *J. Pers. Soc. Psychol.* 90, 999–1013.
13 <http://dx.doi.org/10.1037/0022-3514.90.6.999>
- 14 Juslin, P. N., Laukka, P., 2003. Communication of emotions in vocal expression and music
15 performance: different channels, same code? *Psychol. Bull.* 129, 770–814.
16 <http://dx.doi.org/10.1037/0033-2909.129.5.770>
- 17 Keltner, D., Kring, A.M., 1998. Emotion, social function and psychopathology. *Rev. General*
18 *Psychol.* 2, 320–342. <http://dx.doi.org/10.1037/1089-2680.2.3.320>
- 19 Keltner, D., Tracy, J., Sauter, D. A., Cordaro, D. C., McNeil, G. 2016. Expression of emotion,
20 in: Feldman Barrett, L., Lewis, M., Haviland-Jones, J.M. (Eds.), *Handbook of emotions*.
21 Guilford Press, New York, pp. 467–482.
- 22 Klasen, M., Kreifelts, B., Chen, Y. H., Seubert, J., Mathiak, K., 2014. Neural processing of
23 emotion in multimodal settings. *Front. Hum. Neurosci.* 8, 822
24 <http://dx.doi.org/10.3389/fnhum.2014.00822>
- 25 Knight, E. L., Sarkar, A., Prasad, S., Mehta, P. H., 2020. Beyond the challenge hypothesis: The
26 emergence of the dual-hormone hypothesis and recommendations for future research. *Horm.*
27 *Behav.* in press. <https://doi.org/10.1016/j.yhbeh.2019.104657>
- 28 Kohler, C. G., Walker, J. B., Martin, E. A., Healey, K. M., Moberg, P. J., 2009. Facial emotion
29 perception in schizophrenia: a meta-analytic review. *Schizophr. Bull.* 36, 1009–1019.
30 <http://dx.doi.org/10.1093/schbul/sbn192>
- 31 Kordsmeyer, T. L., Lohöfener, M., Penke, L., 2018. Male facial attractiveness, dominance, and
32 health and the interaction between cortisol and testosterone. *Adapt. Human Behav. Physiol.*, 1–
33 12. <https://doi.org/10.1007/s40750-018-0098-z>
- 34 Kosonogov, V., Titova, A., 2018. Recognition of all basic emotions varies in accuracy and
35 reaction time: A new verbal method of measurement. *Int. J. Psychol.* 16,
36 <http://dx.doi.org/10.1002/ijop.12512>

1 Kraus, M. W., 2017. Voice-only communication enhances empathic accuracy. *Am. Psychol.*
2 72, 644–654. <http://dx.doi.org/10.1037/amp0000147>

3 Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., Wildgruber, D., 2007. Audiovisual integration
4 of emotional signals in voice and face: An event-related fMRI study. *NeuroImage* 37, 1445–
5 1456. <https://doi.org/10.1016/j.neuroimage.2007.06.020>.

6 Krumhuber, E.G., Skora, L., Kuester, D., Fou, L., 2017. A review of dynamic datasets for facial
7 expression research. *Emotion Review* 9, 280–292.
8 <http://dx.doi.org/10.1177/1754073916670022>

9 Krumhuber, E.G., Kappas, A., Manstead, A.S.R., 2013. Effects of dynamic aspects of facial
10 expressions: A review. *Emotion Review* 5, 41–46. <https://doi.org/10.1177/1754073912451349>

11 Lambrecht, L., Kreifelts, B., Wildgruber, D (2014). Gender differences in emotion recognition:
12 impact of sensory modality and emotional category. *Cogn. Emot.* 28, 452–469.
13 <http://dx.doi.org/10.1080/02699931.2013.837378>

14 Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., van Knippenberg, Ad,
15 2010. Presentation and validation of the Radboud Faces Database. *Cogn. Emot.* 24, 1377–1388.
16 <http://dx.doi.org/10.1080/02699930903485076>

17 Lausen, A., Hammerschmidt, K. 2020. Emotion recognition and confidence ratings predicted
18 by vocal stimulus type and acoustic parameters. <https://doi.org/10.31234/osf.io/kqy2n>

19 Lausen, A., Schacht, A., 2018. Gender differences in the recognition of vocal emotions. *Front.*
20 *Psychol.* 9, 882. <https://doi.org/10.3389/fpsyg.2018.00882>

21 Liening, S. H., Josephs, R. A., 2010. It is not just about testosterone: physiological mediators
22 and moderators of testosterone's behavioral effects. *Soc. Personal. Psychol. Compass* 4, 982–
23 994. <http://dx.doi.org/10.1111/j.1751-9004.2010.00316.x>

24 Marsh, A.A., Blair, R.J.R. (2008). Deficits in facial affect recognition among antisocial
25 populations: a meta-analysis. *Neurosci. Biobehav. Rev.* 32, 454–465.
26 <https://doi.org/10.1016/j.neubiorev.2007.08.003>

27 Matas D, Koren L. 2016. A simple method for measuring long-term integrated testosterone
28 levels in men. *PeerJ Preprints* 4:e2069v2. <https://doi.org/10.7287/peerj.preprints.2069v2>

29 Mayer, J. D., Caruso, D. R., Salovey, P., 2016. The ability model of emotional intelligence:
30 Principles and updates. *Emot. Rev.* 8, 290– 300. <http://dx.doi.org/10.1177/1754073916639667>

31 Mazur, A., Booth, A., 2014. Testosterone is related to deviance in male army veterans, but
32 relationships are not moderated by cortisol. *Biol. Psychol.* 96, 72–76.
33 <https://doi.org/10.1016/j.biopsycho.2013.11.015>

34 McEwen, B. S., 2008. Central effects of stress hormones in health and disease: understanding
35 the protective and damaging effects of stress and stress mediators. *Eur. J. Pharmacol.* 583,
36 174–185. <https://doi.org/10.1016/j.ejphar.2007.11.071>

- 1 Mehta, P.H., Prasad, S., 2015. The dual-hormone hypothesis: A brief review and future research
2 agenda. *Curr. Opin. Behav. Sci.* 3, 163–168. <https://doi.org/10.1016/j.cobeha.2015.04.008>
- 3 Mehta, P.H., Welker, K. M., Zilioli, S., Carré, J. M., 2015a. Testosterone and cortisol jointly
4 modulate risk-taking. *Psychoneuroendocrinology* 56, 88–99.
5 <http://dx.doi.org/10.1016/j.psyneuen.2015.02.023>
- 6 Mehta, P. H., Shira, M., Yap, A., J., Prasad, S., 2015b. Dual-Hormone Changes Are Related to
7 Bargaining Performance. *Psychol. Sci.* 26, 866–876.
8 <http://dx.doi.org/10.1177/0956797615602271>
- 9 Mehta, P.H., Josephs, R.A., 2010. Testosterone and cortisol jointly regulate dominance:
10 Evidence for a dual-hormone hypothesis. *Horm. Behav.* 58, 898–906.
11 <https://doi.org/10.1016/j.yhbeh.2010.08.020>
- 12 Miller, R., Plessow, F., Rauh, M., Groschl, M., Kirschbaum, C. (2013). Comparison of salivary
13 cortisol as measured by different immunoassays and tandem mass spectrometry.
14 *Psychoneuroendocrinology* 38, 50–57. <https://doi.org/10.1016/j.psyneuen.2012.04.019>
- 15 Montagne, B., Kessels, R.P., de Haan, E.H.F., Perrett, D.I., 2007. The emotion recognition task:
16 a paradigm to measure the perception of facial emotional expressions at different intensities.
17 *Percept. Mot. Skills* 104, 589–598. <https://doi.org/10.2466/pms.104.2.589-598>
- 18 Nadler, A., Camerer, C. F., Zava, D. T., Ortiz, T. L., Watson, N. V., Carré, J. M, Nave, G.,
19 2019. Does testosterone impair men's cognitive empathy? Evidence from two large-scale
20 randomized controlled trials. *Proc. R. Soc. B* 286: 20191062.
21 <http://dx.doi.org/10.1098/rspb.2019.1062>
- 22 Nitschke, J. P., Bartz, J. A., 2020. Lower digit ratio and higher endogenous testosterone are
23 associated with lower empathic accuracy. *Horm. Behav.* 119, 1–8.
24 <https://doi.org/10.1016/j.yhbeh.2019.104648>
- 25 Nummenmaa, L., Calvo, M. G., 2015. Dissociation between recognition and detection
26 advantage for facial expressions: a meta-analysis. *Emotion* 15, 243–256.
27 <http://dx.doi.org/10.1037/emo0000042>
- 28 Olderbak, S., Wilhelm, O., 2017. Emotion perception and empathy: An individual differences
29 test of relations. *Emotion* 17, 1092–1106. <http://dx.doi.org/10.1037/emo0000308>
- 30 Palermo, R. Coltheart, M., 2004. Photographs of facial expression: accuracy, response times,
31 and ratings of intensity. *Behav. Res. Methods Instrum. Comput.* 36, 634–638.
32 <https://doi.org/10.3758/BF03206544>
- 33 Paulmann, S., Pell, M. D., 2011. Is there an advantage for recognizing multi-modal emotional
34 stimuli? *Motiv. Emot.* 35, 192–201. <https://doi.org/10.1007/s11031-011-9206-0>
- 35 Paulmann, S., Kotz, S.A., 2008. An ERP investigation on the temporal dynamics of emotional
36 prosody and emotional semantics in pseudo- and lexical sentence context. *Brain Lang.* 105, 59–
37 69. <http://dx.doi.org/10.1016/j.bandl.2007.11.005>

1 Paulmann, S., Pell, M.D., Kotz, S.A., 2008. How aging affects the recognition of emotional
2 speech, *Brain Lang.* 104, 262–269. <http://dx.doi.org/10.1016/j.bandl.2007.03.002>

3 Pell, M.D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., Rigoulot, S., 2015. Preferential
4 decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biol.*
5 *Psychol.* 111, 14–25. <http://doi.org/10.1016/j.biopsycho.2015.08.008>

6 Pichon, S., de Gelder, B., Grèzes, J., 2008. Emotional modulation of visual and motor areas by
7 dynamic body expressions of anger. *Soc. Neurosci.* 3, 199–212.
8 <http://dx.doi.org/10.1080/17470910701394368>

9 Piwek, L., Pollick, F., Petrini, K., 2015. Audiovisual integration of emotional signals from
10 others' social interactions. *Front. Psychol.* 6, 611. <http://dx.doi.org/10.3389/fpsyg.2015.00611>

11 Ponzi, D., Zilioli, S., Mehta, P. H., Maslov, A., Watson, N. V., 2016. Social network centrality
12 and hormones: The interaction of testosterone and cortisol. *Psychoneuroendocrinology* 68, 6–
13 13. <http://dx.doi.org/10.1016/j.psyneuen.2016.02.014>

14 Popma, A., Vermeiren, R., Geluk, C.A., Rinne, T., van den Brink, W., Knol, D.L, et al., 2007.
15 Cortisol moderates the relationship between testosterone and aggression in delinquent male
16 adolescents. *Biol. Psychiatry* 61, 405–411. <http://dx.doi.org/10.1016/j.biopsycho.2006.06.006>

17 Prasad, S., Lassetter, B., Welker, K. M., Mehta, P.H., (2019). Unstable correspondence between
18 salivary testosterone measured with enzyme immunoassays and tandem mass spectrometry.
19 *Psychoneuroendocrinology* 109, 104373. <https://doi.org/10.1016/j.psyneuen.2019.104373>

20 R Core Team, 2017. R: A Language and Environment for Statistical Computing. Vienna: R
21 Foundation for Statistical Computing. Available online at: [https:// www.r-project.org](https://www.r-project.org)

22 Recio, G., Sommer, W., Schacht, A., 2011. Electrophysiological correlates of perceiving and
23 evaluating static and dynamic facial emotional expressions. *Brain Res.* 1376, 66–75.
24 <https://doi.org/10.1016/j.brainres.2010.12.041>

25 Rigoulot S, Wassiliwizky, E., Pell, M. D., 2013. Feeling backwards? How temporal order in
26 speech affects the time course of vocal emotion recognition. *Front. Psychol.* 4:367.
27 <https://doi.org/10.3389/fpsyg.2013.00367>

28 Roelofs, K., Bakvisa, P., Hermans, E. J., van Pelt, J., van Honk, J., 2007. The effects of social
29 stress and cortisol responses on the preconscious selective attention to social threat. *Biol.*
30 *Psychol.* 75, 1–7. <https://doi.org/10.1016/j.biopsycho.2006.09.002>

31 Ronay, R., Carney, D. R., 2013. Testosterone's negative relationship with empathic accuracy
32 and perceived leadership ability. *Soc. Psychol. Personal. Sci.* 00, 1–8.
33 <https://doi.org/10.1177/1948550612442395>

34 Roy, A.R.K., Cook, T., Carré, J.M., Welker, K.M., 2019. Dual-hormone regulation of
35 psychopathy: evidence from mass spectrometry. *Psychoneuroendocrinology* 99, 243–250.
36 <https://doi.org/10.1016/j.psyneuen.2018.09.006>

- 1 Rukavina, S., Sachsenweger, F., Jerg-Bretzke, L., Daucher, A. E., Traue, H. C., Walter, S.,
2 Hoffmann, H., 2018. Testosterone and its influence on emotion recognition in young, healthy
3 males. *Psychology* 9, 1814–1827. <http://dx.doi.org/10.4236/psych.2018.97106>
- 4 Schaerlaeken S., Grandjean, D., 2018. Unfolding and dynamics of affect bursts decoding in
5 humans. *PLoS ONE* 13: e0206216. <https://doi.org/10.1371/journal.pone.0206216>
- 6 Sarkar, A., Mehta, P. H., Josephs, R. A., 2019. The dual-hormone approach to dominance and
7 status-seeking, in: Schultheiss O. C., Mehta P. H. (Eds.), *Routledge International Handbook of*
8 *Social Neuroendocrinology* (1st Ed.), Routledge, Abingdon, UK, pp. 113–132.
9 <https://www.routledgehandbooks.com/doi/10.4324/9781315200439-8>
- 10 Sherman, G. D., Mehta, P. H., 2020. Stress, cortisol, and social hierarchy. *Current Opinion in*
11 *Psychology* 33, 227–232. <https://doi.org/10.1016/j.copsyc.2019.09.013>
- 12 Sherman, G. D., Lerner, J. S., Josephs, R. A., Renshon, J., Gross, J. J., 2016. The interaction of
13 testosterone and cortisol is associated with attained status in male executives. *J. Pers. Soc.*
14 *Psychol.* 110, 921–929. <https://doi.org/10.1037/pspp0000063>
- 15 Schirmer, A., 2013. Sex differences in emotion, in: Armony, J., Vuilleumier, P. (Eds.), *The*
16 *Cambridge Handbook of Human Affective Neuroscience* (1st Ed.). Cambridge University Press,
17 New York, pp. 591–611.
- 18 Schultheiss, O.C., Dlugash, G., Mehta, P.H., 2019. Hormone measurement in social
19 neuroendocrinology: a comparison of immunoassay and mass spectrometry methods, in:
20 Schultheiss O.C., Mehta P.H. (Eds.), *Routledge International Handbook of Social*
21 *Neuroendocrinology*. Routledge, Abingdon, UK, pp. 26–41.
22 <https://www.routledgehandbooks.com/doi/10.4324/9781315200439-3>
- 23 Schultheiss, O.C., Wiemers, U.S., Wolf, O.T., 2016. Exploring effects of hydrocortisone on
24 implicit motivation and activity inhibition: a randomized placebo-controlled study. *Adapt.*
25 *Hum. Behav. Physiol.* 2, 267–280. <https://doi.org/10.1007/s40750-016-0043-y>
- 26 Schultheiss, O.C., Schiepe, A., Rawolle, M., 2012. Hormone assays, in: Cooper, H., Camic, P.
27 M., Long, D.L., Panter, A.T., Rindskopf, D., Sher, K.J. (Eds.), *Handbook of research methods*
28 *in psychology*. American Psychological Association, Washington, pp. 489–500.
- 29 Schultheiss, O.C., Stanton, S.J., 2009, in: Harmon-Jones, E. & Beer, J. S. (Eds.), *Methods in*
30 *social neuroscience*. Guilford Press: New York, pp. 17–44.
- 31 Sokolowski, K., Schmalt, H.D., Langens, T.A., Puca, R.M., 2000. Assessing achievement,
32 affiliation, and power motives all at once: The Multi-Motive Grid (MMG). *J. Pers. Assess.* 74,
33 126–145.
- 34 Soto, J.A., Levenson, R.W., 2009. Emotion recognition across cultures: the influence of
35 ethnicity on empathic accuracy and physiological linkage. *Emotion* 9, 874–884.
36 <http://dx.doi.org/10.1037/a0017399>

- 1 Tackett, J. L., Reardon, K. W., Herzhoff, K., Page-Gould, E., Harden, K. P., Josephs, R.A.,
2 2015. Estradiol and cortisol interactions in youth externalizing psychopathology.
3 *Psychoneuroendocrinology* 55, 146 – 153. <http://dx.doi.org/10.1016/j.psyneuen.2015.02.014>
- 4 Taylor, A.E., Keevil, B., Huhtaniemi, I.T. (2015). Mass spectrometry and immunoassay: how
5 to measure steroid hormones today and tomorrow. *Eur. J. Endocrinol* 173, D1- 12.
6 <http://dx.doi.org/10.1530/EJE-15-0338>
- 7 Thompson, A. E., Voyer, D., 2014. Sex differences in the ability to recognize non-verbal
8 displays of emotion: A meta-analysis. *Cogn. Emot.* 28, 1164–1195.
9 <http://dx.doi.org/10.1080/02699931.2013.875889>
- 10 Tops, M., Boksem, M. A., 2010. Cortisol involvement in mechanisms of behavioral inhibition.
11 *Psychophysiology* 48, 723–732. <http://dx.doi.org/10.1111/j.1469-8986.2010.01131.x>
- 12 van Bokhoven, I., van Goozen, S.H., van Engeland, H., Schaal, B., Arseneault, L., et al., 2005.
13 Salivary cortisol and aggression in a population-based longitudinal study of adolescent males.
14 *J. Neural Transm.* 112, 1083–1096. <http://dx.doi.org/10.1007/s00702-004-0253-5>
- 15 van Honk, J., Schutter, D. J., Bos, P. A., Kruijt, A. W., Lentjes, E. G., Baron-Cohen, S., 2011.
16 Testosterone administration impairs cognitive empathy in women depending on second-to-
17 fourth digit ratio. *Proc. Natl. Acad. Sci.* 108, 3448–3452.
18 <https://doi.org/10.1073/pnas.1011891108>
- 19 van Honk, J., Schutter, D.J., 2007. Testosterone reduces conscious detection of signals serving
20 social correction: Implications for antisocial behavior. *Psychol. Sci.* 18, 663–667.
21 <https://doi.org/10.1111/j.1467-9280.2007.01955.x>
- 22 van Honk, J., Tuiten, A., van den Hout, M., Koppeschaar, H., Thijssen, J., de Haan, E., et al.,
23 2000. Conscious and preconscious selective attention to social threat: Different neuroendocrine
24 response patterns. *Psychoneuroendocrinology* 25, 577–591. [https://doi.org/10.1016/S0306-](https://doi.org/10.1016/S0306-4530(00)00011-1)
25 [4530\(00\)00011-1](https://doi.org/10.1016/S0306-4530(00)00011-1)
- 26 van Honk, J., Tuiten, A., Verbaten, R., van den Hout, M., Koppeschaar, H., Thijssen, J., de
27 Haan, E., 1999. Correlations among salivary testosterone, mood, and selective attention to
28 threat in humans. *Horm. Behav.* 36, 17–24. <https://doi.org/10.1006/hbeh.1999.1521>
- 29 van Honk, J., Tuiten, A., van den Hout, M., Koppeschaar, H., Thijssen, J., de Haan, E., et al.,
30 1998. Baseline salivary cortisol levels and preconscious selective attention for threat: A pilot
31 study. *Psychoneuroendocrinology* 23, 741–747. [https://doi.org/10.1016/S0306-](https://doi.org/10.1016/S0306-4530(98)00047-X)
32 [4530\(98\)00047-X](https://doi.org/10.1016/S0306-4530(98)00047-X)
- 33 Verduyn, P., Delvaux, E., van Coillie, H., Tuerlinckx, F., van Mechelen, I., 2009. Predicting
34 the duration of emotional experience: two experience sampling studies. *Emotion* 9, 83–91.
35 <https://doi.org/10.1037/a0014610>
- 36 Viau, V., 2002. Functional cross-talk between the hypothalamic-pituitary-gonadal and- adrenal
37 axes. *J. Neuroendocrinol.*, 14, 506–513. <http://dx.doi.org/10.1046/j.1365-2826.2002.00798.x>

- 1 Vongas, J. G., Hajj, R. A., 2017. The effects of competition and implicit power motive on men's
2 testosterone, emotion recognition, and aggression. *Horm. Behav.* 92, 57–71.
3 <http://dx.doi.org/10.1016/j.yhbeh.2017.04.005>
- 4 Waaramaa, T., 2017. Gender differences in identifying emotions from auditory and visual
5 stimuli. *Logop. Phoniatr. Voco.* 42, 160–166. <https://doi.org/10.1080/14015439.2016.1243725>
- 6 Wagner, H.L., 1993. On measuring performance in category judgement studies of nonverbal
7 behaviour. *J. Nonverbal Behav.* 17, 3–28. <http://dx.doi.org/10.1007/BF00987006>
- 8 Wells L.J., Gillespie S.M., Rotshtein, P., 2016. Identification of emotional facial expressions:
9 effects of expression, intensity, and sex on eye Gaze. *PLoS ONE* 11, e0168307.
10 <https://doi.org/10.1371/journal.pone.0168307>
- 11 Welker, K.M., Prasad, S., Srivastava, S., Mehta, P.H., 2017. Basal cortisol's relation to
12 testosterone changes may not be driven by social challenges. *Psychoneuroendocrinology* 85,
13 1–5. <http://dx.doi.org/10.1016/j.psyneuen.2017.07.493>
- 14 Welker, K.M., Lassetter, B., Brandes, C.M., Prasad, S., Koop, D.R., Mehta, P.H. (2016). A
15 comparison of salivary testosterone measurement using immunoassays and tandem mass
16 spectrometry. *Psychoneuroendocrinology* 71, 180–188.
17 <http://dx.doi.org/10.1016/j.psyneuen.2016.05.022>
- 18 Welker, K.M., Lozoya, E., Campbell, J.A., Neumann, C.S., Carré, J.M., 2014. Testosterone,
19 cortisol, and psychopathic traits in men and women. *Physiol. Behav.* 129, 230–236.
20 <https://doi.org/10.1016/j.physbeh.2014.02.057>
- 21 Wendt, B., Scheich, H., 2002. The “Magdeburger Prosodie Korpus” – a spoken language corpus
22 for fMRI-Studies, in: Bel, B., Marlien, I. (Eds.), *Speech Prosody* (pp.699–701). Aix-en-
23 Provence: SproSIG
- 24 Williams, L.M., Mathersul, D., Palmer, D.M., Gur, R.C., Gur, R.E., Gordon, E., 2009. Explicit
25 identification and implicit recognition of facial emotions: I. Age effects in males and females
26 across 10 decades. *J. Clin. Exp. Neuropsychol.* 31, 257–277.
27 <http://dx.doi.org/10.1080/13803390802255635>
- 28 Wirth, M. M., Schultheiss, O. C., 2007. Basal testosterone moderates responses to anger faces
29 in humans. *Physiol. Behav.* 90, 496–505. <http://dx.doi.org/10.1016/j.physbeh.2006.10.016>
- 30 World Medical Association, 2013. World Medical Association Declaration of Helsinki: ethical
31 principles form medical research involving human subjects. *JAMA* 310, 2191–2194.
32 <http://dx.doi.org/10.1001/jama.2013.281053>
- 33 Zilioli, S., Bird, B. M., 2017. Functional significance of men's testosterone reactivity to social
34 stimuli. *Front. Neuroendocrinol.* 33, 1–18. <http://dx.doi.org/10.1016/j.yfrne.2017.06.002>
- 35 Zilioli, S., Ponzi, D., Henry, A., Maestripieri, D., 2015. Testosterone, cortisol and empathy:
36 evidence for the dual-hormone hypothesis. *Adapt. Hum. Behav. Physiol.* 1, 421 – 433.
37 <https://doi.org/10.1007/s40750-014-0017-x>

- 1 Zilioli, S., Caldbick, E., Watson, N.V., 2014. Testosterone reactivity to facial display of
- 2 emotions in men and women. *Horm. Behav.* 65, 461–468.
- 3 <https://doi.org/10.1016/j.yhbeh.2014.04.006>

Acknowledgments

The authors thank Edmund Henniges for technical support, Saskia Brueckner, Marc Koehler and Isabel Noethen for help with data acquisition and all individuals who participated in the research presented here. Additionally, we thank Pranjali Mehta and our anonymous reviewers for feedback on earlier versions of this manuscript.

Author contributions

A. L. designed the research with input from C. B., L. P. and A.S.; C. B. collected part of the data and wrote on the method part; A.L. analysed the data and wrote the paper with input from C. B., L. P. and A.S.

Funding

This research was funded by Leibniz ScienceCampus “Primate Cognition”– Project number 6900199 Leibniz-WissenschaftsCampus and Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project number 254142454/GRK 2070.

Competing Interests

The authors declare no competing interests.

Data availability

The dataset generated and analysed for the current study is available at osf.io/2ayms.